RESEARCH ARTICLE

A role for verb regularity in the L2 processing of the Spanish subjunctive mood: Evidence from eye-tracking

Sara Fernández Cuenca[1]* and Jill Jegerski[2]

[1]Wake Forest University, Winston-Salem, NC, USA; [2]University of Illinois at Urbana-Champaign, Urbana, IL, USA
*Corresponding author. Email: fernans@wfu.edu

Abstract

The present study investigated the second language processing of grammatical mood in Spanish. Eye-movement data from a group of advanced proficiency second language users revealed nativelike processing with irregular verb stimuli but not with regular verb stimuli. A comparison group of native speakers showed the expected effect with both types of stimuli, but these were slightly more robust with irregular verbs than with regular verbs. We propose that the role of verb form regularity was due to the greater visual salience of Spanish subjunctive forms with irregular verbs versus regular verbs and possibly also due to less efficient processing of rule-based regular inflectional morphology versus whole irregular word forms. In any case, the results suggest that what appeared to be difficulty with sentence processing could be traced back to word-level processes, which appeared to be the primary area of difficulty. This outcome seems to go against theories that suggest that L2 sentence processing is shallow.

Introduction

The question of whether adult second language (L2) learners can be similar to native (L1) speakers is fundamental in the study of adult second language acquisition, including research on nonnative sentence processing. The last decade has seen the emergence of different theories that seek to explain how sentences are processed in a nonnative language, with particular attention to identifying areas of difficulty. For one, Clahsen and Felser (2006, 2018) have suggested that "structural processing is compromised in nonnative comprehension" (Felser & Cunnings, 2012, p. 600) and that an intrinsic deficit in syntactic and morphosyntactic representations and a failure to integrate these during real-time language comprehension can best characterize L2 online sentence comprehension. In a closely related theory, Cunnings (2017) has proposed that the primary limitations lie not in syntax and morphosyntax, but in a

greater susceptibility to interference during retrieval from memory, although the empirical predictions of the two theories are often the same. However, the Lexical Bottleneck Hypothesis (Hopp, 2014, 2018) proposes that online effects that appear to indicate an issue with syntax and morphosyntax in sentence processing can in reality be indirect effects of difficulty with lexical access (Hopp, 2017a).

Research related to the newer theory of Hopp (2014, 2018) examines factors at the word level and how they might have implications for processing at the phrase and sentence level, given that lexical access and the representation of many word-level details (e.g., word class, verb subcategorization, number) logically precedes the processing of syntax and morphosyntax. In one example, lower verb frequency was associated with slower processing of cleft sentence structure (e.g., *It was Andrew who forced Piper to steal the camera.*) among L2 learners but not native speakers (Hopp, 2016). Less exposure to lower frequency words is generally associated with delayed lexical access in the processing of individual words (e.g., Brysbaert et al., 2018), an effect that has been widely observed with L1 language users in work on lexical processing and that can be more pronounced with L2 learners than with native speakers (e.g., Cop et al., 2015). Hopp's (2016) study showed that such frequency-related processing difficulty at the word level can delay higher level structural processes for establishing sentence word order, so an apparent struggle with syntactic processing among L2 learners can be traced back to lexical processing. Other word level factors that have been shown to affect syntactic and morphosyntactic processing include the cognate status of target words between participants' two languages (Hopp, 2017b; Miller, 2014) and knowledge of individual lexical items such as gender assignment for nouns (Hopp, 2013; Lemhöfer et al., 2014).

Like this previous work, the present study also examined the role of a lower-level factor (form regularity of Spanish verbs) in the processing of a higher-level morphosyntactic phenomenon (the subjunctive mood in embedded clauses) with the goal of determining whether any observed difficulty in the processing of mood might be better accounted for in terms of generalized difficulty processing phrases and sentences (Clahsen & Felser, 2006, 2018) or as indirect effects of difficulty at the level of individual words (Hopp, 2014, 2018). The word-level factor of interest, verb form regularity, has been said to facilitate the acquisition and use of the subjunctive among L2 learners (Collentine, 1997). Such a role for morphological regularity might be explained in terms of perceptual salience (change in the verb stem vs. a single vowel) or item-specific lexical information for verb conjugation class (*-ar, -er, -ir*), as will be explained in the background sections that follows on the Spanish subjunctive and its acquisition by L2 learners. In either case, form irregularity would facilitate the processing of a verb so that it is accurately marked for either indicative or subjunctive mood, which must occur prior to the processing of the subjunctive at the morphosyntactic level within the broader sentence. In other words, if the embedded clause verb is more likely to be accurately processed for mood when it is irregular, this would also improve the chances of successful processing of the mood dependency between that verb and the main clause trigger verb (to be discussed in greater detail in the following section). In this sense, verb regularity is analogous to other word-level factors like word frequency and cognate status in that it could indirectly affect processing at the phrase level.

The subjunctive mood in Spanish

Grammatical mood is a type of inflectional morphology on verbs that communicates semantic modality, or the speaker's attitude toward the propositional content of a

phrase (Bosque, 2012). In Spanish, finite verbs are inflected for one of three moods: indicative, which represents the unmarked default for assertions; imperative, which is used for direct commands; and subjunctive, which is used to convey attitudes such as doubt, desire, and conjecture. The subjunctive mood typically appears in an embedded clause of a complex sentence and is selected by the lexical semantics of a verb or other expression in the matrix clause. For example, a main clause like *Dudo que…* "I doubt that…" or *Espero que…* "I hope that…" would trigger subjunctive morphology on the verb in the embedded clause because the speaker is expressing doubt or desire.

In terms of morphology, finite verbs in Spanish contain a stem and one to three of the following inflectional suffixes: a thematic vowel that indicates to which of the three conjugation classes the verb belongs (i.e., *-ar, -er, -ir*); a composite morpheme for tense, aspect, and mood; and an agreement morpheme for person and number (ibid.). In the present tense, which was used in the eye-tracking stimuli for the present study, the subjunctive mood is not marked with the addition of a suffix but rather by switching the thematic vowel, so *a* changes to *e* and *e* and *i* change to *a*. Hence, the vowel does not usually indicate mood because two of the three vowels, *a* and *e*, which are also by far the most common, are used to mark both indicative and subjunctive, depending on the verb. The only way to know which is which is to compare the thematic vowel with the conjugation class of the individual verb, which means accessing this item-specific information in the lexicon. Moreover, for regular verbs in the present, the vowel switch is the only difference between the indicative and subjunctive, so mood is not very salient to the reader or listener (e.g., *escuchan*$_{\text{IND}}$ vs. *escuchen*$_{\text{SUBJ}}$ "they listen"). Irregular verbs, however, show a stem change in addition to the switch of thematic vowel, so the difference is more readily apparent (e.g., *tienen*$_{\text{IND}}$ vs. *tengan*$_{\text{SUBJ}}$ "they have").

Thus, the Spanish subjunctive is semantically abstract and also linguistically complex because it involves verb morphology, sentence-level semantics, and morphosyntax. Indeed, there is an extensive body of literature in theoretical linguistics that attempts to explain the subjunctive and its many nuances (see ibid., for an overview), which include obligatory use in some linguistic contexts and variable use in others (Gudmestad, 2010, 2012a; Poplack et al., 2018). Not surprisingly, mood tends to be acquired later by children than many other aspects of grammar and different contexts of use appear at different ages. The morphological inflections for mood are first seen with direct commands and are typically acquired by around age 2 (López Ornat et al., 1994). The subjunctive appears next with temporal adverbials (e.g., *cuando* "when…," *antes de que* "before…") and in sentential complement clauses with volitional predicates (like those used in the stimuli for the present study), both of which show mostly adult-like use by around age 4 (Blake, 1983; Sánchez-Naranjo & Pérez-Leroux, 2010). Other types of sentential complement clauses are acquired over the next several years; for example, by around age 9 for predicates of doubt, attitude, and assertion (Blake, 1983). Moreover, in addition to linguistic development, cognitive development appears to be a factor in learning uses of the subjunctive mood that rely on epistemic aspects of semantics, as these rely on the capacity to understand false beliefs (Pérez-Leroux, 1998).

From the perspective of sentence processing, the morphosyntactic relationship between a trigger expression in the matrix clause and the subjunctive morphology on the verb in an embedded clause could be classified as a *distance dependency*, a phenomenon that is of particular interest in the study of nonnative processing because of the high processing demands it incurs (Clahsen & Felser, 2006). To our knowledge, there are no published eye-tracking studies of the L1 processing of Spanish mood. There is one published study using self-paced reading (Demestre & García-Albea, 2004, Experiment 2), but the target was syntactic ambiguity rather than

morphosyntax per se. The stimulus sentences were also quite different from those for the present study, as the subjunctive appeared in the past tense in very complex triclausal sentences that were always grammatical, but in the present study it appeared in the present tense in biclausal sentences that varied with regard to the grammaticality of the critical verb. Still, the results are broadly relevant to the present study in that they showed that experimental manipulations of mood can bring about online effects, at least with native speakers.

The Spanish subjunctive in adult second language acquisition

The acquisition of the Spanish subjunctive mood by adult learners has been the object of a fair amount of empirical investigation, probably in part because of its notoriety among students and instructors in the context of formal language instruction (see Collentine, 2014, for a comprehensive review). As outlined in the previous section, grammatical mood involves semantic abstraction and linguistic complexity, which makes it difficult to acquire, plus mood morphemes typically involve the switching of a single vowel, so they are often nonsalient. Furthermore, subjunctive morphology is usually redundant and thus not critical to meaningful communication (Lee, 1987; Terrell et al., 1987), which in turn makes the form more difficult to acquire (Leow, 1993; VanPatten, 1994, 1996). To illustrate, in a sentence like *Espero que escuchen*SUBJ "I hope they listen," both the matrix clause verb *espero* "hope" and the inflectional morphology on the embedded clause verb *escuchen*SUBJ "listen" convey the speaker's attitude of wishing for the event to happen.

Despite these obstacles, research has suggested that some uses of the subjunctive can be acquired to some degree. There is evidence that oral production improves with immersion experience (Isabelli & Nishida, 2005; Lubbers Quesada, 1998). Studies in the generative framework have found that performance on interpretation and judgment measures can be high or even nativelike among adult learners with a very high level of L2 proficiency, for trigger contexts such as volitional predicates (Borgonovo et al., 2005; Iverson et al., 2008; Massery, 2009) and negated epistemic and perception predicates (Borgonovo & Prévost, 2003; Iverson et al., 2008). Variationist work has observed that the oral production patterns of high-proficiency L2 learners can closely resemble those of native speakers in terms of frequency and contextual factors that shape variation; the only point of divergence was with the discourse pragmatic variable of hypotheticality (Gudmestad, 2012a).

Research on the L2 acquisition of the Spanish subjunctive has also observed that empirical findings can vary according to the experimental task or measure. First, more targetlike production seems to occur with a greater focus on form: in written production versus an oral interview (Terrell et al., 1987), in a controlled production task versus an oral interview (Collentine, 1995), and in a verb elicitation task more than in a clause elicitation task and in both of those more than in a role play (Gudmestad, 2012a). Second, written comprehension might favor more nativelike subjunctive use as compared to oral production (Geeslin & Gudmestad, 2008; Montrul, 2011). Task effects can also interact with other variables, for example, verb form regularity (Geeslin & Gudmestad, 2008; Gudmestad, 2012b). Finally, performance on a number of untimed and largely form-focused written measures has been found to correlate with general metalinguistic knowledge of Spanish (Correa, 2011) and with explicit knowledge of Spanish mood (Gutiérrez, 2017), whereas there was no correlation of explicit knowledge with accuracy in an oral interview.

Thus, the choice of investigative method appears to be important, with untimed and form-focused tasks being most affected by explicit and metalinguistic knowledge. The same participants who perform quite well on untimed written assessments can struggle with a more authentic communicative task like an oral interview. However, there is evidence that even native speakers might use the subjunctive less in oral production than in writing (Geeslin & Gudmestad, 2008), so oral production also has limitations as an experimental measure. Another way to get at this issue is with a real-time measure of language comprehension such as self-paced reading or eye tracking, which have relatively realistic time constraints and the potential for focus on meaning over form (Keating & Jegerski, 2015), yet they also allow the researcher to target a very specific linguistic form using controlled written stimuli. To our knowledge, only one previous study has taken this approach.

Cameron (2011, 2017) employed the self-paced reading method to study L2 comprehension of the subjunctive with impersonal expressions of certainty such as *Es probable que…* "It's likely that…" and *Es cierto que…* "It's true that…." The experimental task was focused on the meaning of the sentence; a picture was displayed for each stimulus sentence and participants were asked to indicate whether it matched the meaning of the sentence. A comparison group of native speakers showed longer reading times following the critical verb in the ungrammatical condition versus in the grammatical condition, but no reading time differences according to whether the picture matched the stimulus sentence. Conversely, three groups of L2 learners at different proficiency levels all showed sensitivity to the match between the sentence and the picture, but not to the grammaticality of mood in the sentence. Cameron concluded that L2 learners do not process the subjunctive like native Spanish users, in line with the Shallow Structure Hypothesis (Clahsen & Felser, 2006, 2018). Nevertheless, an important limitation of this study was that it only included regular verbs, even though there is evidence that irregular verbs may provide an advantage to L2 learners with the subjunctive (Collentine, 1997; Gudmestad, 2012b).

The role of verb regularity in the acquisition of the Spanish subjunctive was first investigated by Collentine (1997), who suggested that verbs with irregular stems were more likely to be noticed and subsequently acquired by learners because the mood contrast is more salient with irregular verbs than with regular ones, in which a single vowel changes to mark mood. Indeed, the participants in Collentine's experiment took more time to respond to irregular verb items than to regular ones in a meaning-oriented scrambled sentence task. A number of studies with a variationist approach have observed that L2 learners of Spanish use the subjunctive more with irregular verbs than with regular ones in oral interviews (Lubbers-Quesada, 1998), in a written binary choice paragraph completion task (Gudmestad, 2006), and in three different oral elicitation tasks (Gudmestad, 2012a). Still, there is also evidence that such effects do not appear on all experimental measures or at all levels of proficiency (Gudmestad, 2012b), and that the reverse effect can even occur in some contexts (Geeslin & Gudmestad, 2008). A more recent study by Gallego and Pozzi (2018) found that irregular morphology influenced recognition and production of the subjunctive in both the aural and written modality, but at different rates depending on the task.

In sum, the Spanish subjunctive has received a good amount of attention in SLA research and there is evidence that the choice of experimental task is important. Nevertheless, only one prior investigation employed a real-time processing measure, self-paced reading, and it employed stimulus sentences with only regular verbs, which have been associated with less nativelike performance in studies using other research methods. In the present study, we addressed these limitations by including stimuli with

both regular and irregular verbs. In addition, this study also employed eye tracking, a more nuanced measure of sentence processing than self-paced reading. More specifically, because eye tracking provides data on both early processing and later stages, it can determine whether L2 readers might be slower to integrate different types of linguistic information during processing than L1 readers (e.g., Felser et al., 2012). Another advantage of eye tracking is that it has greater ecological validity as an experimental measure of reading than self-paced reading: rereading is possible in eye tracking but not in self-paced reading, sentences must be displayed as individual words or phrases in self-paced reading but not with eye tracking, and participants have to press a button after each word or phrase during self-paced reading, but not with eye tracking.

Given this background, the present study posited the following two research questions:

1. Do advanced L2 readers show online sensitivity to Spanish mood while reading sentences for comprehension? How do they compare to native readers in this regard?
2. Does the regularity of verbs marked for mood affect online sensitivity among advanced L2 readers? How do they compare to native readers in this regard?

These two research questions were based on current theoretical debates in L2 sentence processing, as outlined in the "Introduction" section of this article. More specifically, these questions were posed to determine whether difficulty in processing Spanish mood (if evident) is caused by generalized difficulty with the syntax and morphosyntax of the form, as proposed by the Shallow Structure Hypothesis (Clahsen & Felser, 2006, 2018), or alternatively, such difficulty is an indirect effect of lower level factors that affect the processing of individual lexical items, as proposed by the Lexical Bottleneck Hypothesis (Hopp, 2018). Results that reflect a lack of online sensitivity to mood in all contexts, regardless of verb regularity, would support theoretical claims of difficulty with syntax and morphosyntax (Clahsen & Felser, 2006, 2018), whereas a role for verb regularity would support the claim that the primary difficulty arises at the word level (Hopp, 2018).[1]

---

[1]It should be noted that the Shallow Structure Hypothesis proposes an increased role for lexical information as a way of compensating for difficulty in the L2 processing syntax and morphosyntax and, on the surface, this claim may sound like it would also be supported by a role for verb regularity in the processing of the subjunctive because word form information is stored in the lexicon. However, the Shallow Structure Hypothesis juxtaposes the lexicon and syntax/morphosyntax in this manner when the two are alternate cues in the interpretation of a single detail of sentence processing (in either a competing or complementary manner). To illustrate, in the processing of ambiguous relative clauses (Felser et al., 2003; Papadopoulou & Clahsen, 2003), the "lexical bias" (Felser et al., 2003, p. 457) of the preposition "with," which leads to low attachment of relative clauses, is contrasted with language-specific phrase structure principles that lead to high attachment because both provide cues in the attachment of relative clauses, even though the former is lexical and the latter syntactic. In the present study, however, verb regularity did not provide any type of cue regarding the mood of a given verb or of the clause or sentence in which it appeared. Verb regularity was not a potential substitute or competitor for the syntactic and morphosyntactic information that is required to process mood because irregular verbs occur in both the subjunctive and the indicative, as do regular verbs. Hence, it appears that verb regularity in the processing of Spanish mood does not fit with the type of lexical information that is of interest in the Shallow Structure Hypothesis and that the theory would therefore not account for a role for verb regularity as an experimental outcome.

Method

Participants

Twenty Spanish native speakers and 20 high-proficiency Spanish L2 speakers were recruited at a large university in the Midwestern United States (see Table 1). The native speakers were raised in Spanish-speaking countries and moved to the United States to pursue college degrees. They did not acquire proficiency in English until after puberty, although some reported minimal exposure to the language in childhood, typically using school class time of 1 hour or less per week. The L2 learners were native speakers of English who learned Spanish after puberty. Proficiency was measured with a modified version of the DELE standardized Spanish proficiency test (Montrul & Slabakova, 2003). Cronbach's alpha (a measure of reliability) for the test was .73, which is slightly lower than in previous research using the same instrument (.83, Montrul & Ionin, 2012; .84, Montrul et al., 2008).

The sample size of 40 total participants was chosen based on common practice in L2 eye-tracking studies, but power analysis for linear mixed-effects models was also conducted using the *simr* package (Green & MacLeod, 2016) in R (R Core Team, 2021), based on 200 simulations per effect, a moderate effect size of .40, 40 total participants, and 32 stimulus items per participant. Power to detect the effects of group, grammaticality, verb regularity, and three interactions of interest, grammaticality × group, grammaticality × verb regularity, and grammaticality × group × verb regularity, in the total dwell time data from the critical region of interest was calculated. Estimates ranged from 79.50% to 100.00% power, with the lowest value corresponding to the three-way interaction.

Materials

The experimental stimuli were complex sentences with a critical verb in an embedded clause that required the subjunctive mood because of a trigger verb in the preceding

Table 1. Language background information

	L1 (n = 20)			Advanced L2 (n = 20)		
	M	SD	range	M	SD	range
Age	30.90	8.08	20–56	31.20	9.80	21–64
Age of acquisition						
English	9.90	3.38	6–15	.15	.67	0–3[a]
Spanish	.15	.48	0–2[a]	13.10	2.50	10–19
DELE scores	48.25	1.51	45–50	45.20	2.87	38–49
Self-ratings-English						
Understanding	8.65	1.26	5–10	9.95	.22	9–10
Speaking	7.75	1.25	5–10	9.90	.30	9–10
Reading	8.75	1.44	5–10	9.95	.22	9–10
Self-ratings-Spanish						
Understanding	9.75	.71	7–10	8.05	.75	7–9
Speaking	9.90	.30	9–10	7.80	1.15	5–9
Reading	9.90	.30	9–10	8.25	1.11	6–10

Note: The maximum score was 50 for the DELE and 10 for self-rated proficiency.
[a]One of the L1 participants reported an Age of Acquisition for Spanish of 1, but the person also reported no other languages in early childhood. One L1 participant reported an age of 2 for Spanish with early exposure to Basque, but also referred to Spanish as their "mother tongue." One L2 participant reported an age of 3 for English with early exposure to Arabic, but also referred to English as their "first language."

main clause.[2] As illustrated in (1) and (2), the critical verb with mood was either regular or irregular. The regular verbs required only a switch of thematic vowel to mark the subjunctive mood (e.g., *comen* → *coman*), whereas the irregular verbs required the vowel switch plus a stem change that included the addition of a voiced velar stop /g/ (e.g., *tienen* → *tengan*). There are two different types of irregularity with Spanish mood (Gudmestad, 2012b) and we chose this type with the additional consonant because it meant that word length was the same for both indicative and subjunctive moods.

(1) *Regular Verb Stimulus*
    a. El ministro espera que los ciudadanos aprueben su propuesta. *Grammatical*
    b. El ministro espera que los ciudadanos aprueban su propuesta. *Ungrammatical*
    "The minister hopes that the citizens approve$_{SUBJ/IND}$ his proposal."

(2) *Irregular Verb Stimulus*
    a. Los supermercados piden que los clientes traigan sus bolsas. *Grammatical*
    b. Los supermercados piden que los clientes traen sus bolsas.   *Ungrammatical*
    "Supermarkets ask that their customers bring$_{SUBJ/IND}$ their bags."

The regular and irregular verbs used in the stimulus sentences were both high frequency (regular: M = 3.40; irregular: M = 3.84 log frequency per million words; Cuetos et al., 2011) but still differed significantly from each other ($t = 2.29$, $p = .01$), which is typically the case because verb irregularity is associated with high frequency (e.g., Pinker, 1999). Because frequency differences can lead to differences in processing speed at the word level that might translate into difficulty in processing morphosyntax at the sentence level (Hopp, 2016; Jegerski & Fernández Cuenca, 2019), word frequency of the critical verbs was included as a covariate in the statistical analyses of the eye-movement data for this study, as will be shown in the "Results" section.

It should also be noted that the regular and irregular verbs differed with regard to the visual salience of the mood distinction because of the nature of Spanish irregular verbs. With regular verbs, the mood distinction is expressed with a change of thematic vowel, a single letter/phoneme, but with irregular verbs, there is a change of thematic vowel plus a stem change, so the indicative and subjunctive forms differ in terms of multiple letters/phonemes. Hence, morphological regularity and form salience cannot be teased apart in Spanish, but both are word-level factors, so this does not affect the fit of the stimulus design with the theoretical framing of the study.

The trigger verbs for the main clause were chosen based on minimal variation in their selection of the subjunctive mood and high enough word frequency to ensure that they would be known to the advanced level L2 participants in this study (within the top 5,000 "core" Spanish words; Davies, 2006). They also had to fit in coherent sentences with the critical verbs and following the stimulus template. A total of 18 different trigger verbs appeared a mean of 1.7 times (range 0–5) in each of the two stimulus sets (regular verb stimuli and irregular verb stimuli), with the result being that 22 of the 32 stimuli for

---

[2]The stimuli did not include the opposite pattern, in which a subjunctive verb appeared in an embedded clause in the absence of a trigger verb in the main clause. The primary reason for this was that subjunctive verb forms are much less frequent than indicative forms, with a distribution of around 10% subjunctive and 90% indicative (Kanwit & Geeslin, 2018). Hence, for the stimulus condition in question, the critical verb in the ungrammatical condition would also be much less frequent than in the grammatical condition and any increase in dwell times would therefore be difficult to interpret. For this reason, the stimuli with subjunctive verbs were always grammatical.

each verb type (68.8%) had identical trigger verbs to the corresponding items in the other set and the remainder had trigger verbs that were the same ones repeated a different number of times (e.g., *pedir* "to request" appeared once in the irregular verb stimuli and twice in the regular verb stimuli; 12.5%) or were different verbs that shared the same critical characteristics of being strong subjunctive triggers and of relatively high word frequency (18.8%). The lists of trigger verbs in the two sets were of similar log frequency per million words, as shown by an independent samples t-test: $t(62) = .02$, $p = .99$ that included a log frequency for every stimulus item, even if the trigger verb was repeated across multiple items.

Given the variation associated with some uses of the subjunctive in Spanish, we normed the stimuli with a different group of 20 native speakers prior to the experiment. Acceptability judgements on a 5-point Likert scale showed a mean rating of 1.00 (1: "Completely unacceptable") for the ungrammatical items and 4.90 (5: "Totally acceptable") for the grammatical items. The mean acceptability score and range of responses for each sentence can be found in the online supplementary materials.

Thirty-two stimuli with regular verbs and 32 with irregular verbs were rotated across four counterbalanced presentation lists such that each contained 16 stimuli with regular verbs and 16 with irregular ones (each with eight grammatical and eight ungrammatical). A total of 13 irregular verbs and 23 regular verbs were included and were used one to three times per list, with a mean of 1.78 stimuli with each irregular verb and 1.38 stimuli with each regular verb in each list. Each sentence appeared only once in either condition per list (i.e., grammatical or ungrammatical). The 32 target stimuli in each list were combined with 32 distractors, stimuli for another experiment on nonlocal verbal number agreement (as illustrated in Example 3), and 64 fillers that were all also 50/50% grammatical/ungrammatical.[3] The fillers were of a visual length consistent with that of the experimental sentences and contained several different types of grammar errors to maximize distraction, including erroneous prepositions, number agreement in the noun phrase, gender agreement, missing complementizers, and definite articles. Two examples are provided in (4) and (5). The 128 total sentences were presented in pseudorandom order such that no two sentences of the same type appeared in succession. All eye-tracking materials and stimulus counterbalancing followed the recommendations of Keating (2014; Keating & Jegerski, 2015). The complete set of experimental stimuli can be found in the online supplementary materials.

(3) *Distractor with Verbal Number Agreement*
  a. El paquete que pidió la secretaria llegó esta tarde a las cinco.   *Grammatical*
  b. Los paquetes que pidió la secretaria llegó esta tarde a las cinco. *Ungrammatical*
  "The package(s) that the secretary requested arrived$_{SING}$ this afternoon at five."

(4) *Filler Sentence with an Erroneous Preposition*
  El veterinario es un profesional que cuida *por los animales.
  "A veterinarian is a professional who cares for animals."

---

[3]In the target stimuli, the critical verb in the embedded clause always appeared in the subjunctive mood in the grammatical condition and in the indicative in the ungrammatical condition. However, this association between subjunctive/grammatical and indicative/ungrammatical was not present in the overall eye-tracking experiment because the distractors and fillers contained many instances of the indicative in both grammatical and ungrammatical sentences, including in embedded clauses. The subjunctive was not used in any of the distractors or fillers, as we wanted to avoid additional exposure to the target form, which is more marked than the default indicative.

(5) *Filler Sentence without an Error*
En El Salvador, el día de las madres se celebra el 10 de mayo.
"In El Salvador, Mother's Day is celebrated on May 10."

Procedure

Written instructions and eight practice trials preceded the experiment, which was run on a desktop mount Eyelink 1000 eye tracker (SR Research, 2005) with chin and forehead rests, a sampling rate of 1000 Hz, and tracking of the right eye only. The participant sat 39 inches from the display, which was presented on a 22-inch ViewSonic monitor. The eye tracker was calibrated with a nine-point grid and validated to ensure a maximum error of .5 degrees. Calibration was conducted before and after the practice trials, after the participant had completed half of the 128 experimental trials, and additionally as needed, based on an automatic drift check that was included at the beginning of each trial. Stimulus sentences and comprehension questions were presented in black, 24-point Tahoma font on a white background, with each stimulus presented as a single line of text. Participants proceeded from one screen to the next using a green button and responded to comprehension questions using buttons marked "A" and "B" on a Microsoft Sidewinder game controller (the standard response device that came with the Eyelink 1000 equipment package).

Participants were told that the test targeted reading comprehension and they answered a meaning-based comprehension question after each stimulus. Participants were offered breaks from reading after the practice and after they had read half of the 128 sentences. After the reading task, participants completed a language background questionnaire, the proficiency test, and a debriefing questionnaire. Most of the participants for this study also participated in a separate research session for another experiment, during which they read a second list of the same type of stimuli (the other 16 from the 32 total of each type), but no participant ever read the same stimulus sentence more than once in any version. The other experiment examined the role of task goals in sentence processing, specifically reading for comprehension versus reading to judge acceptability, and the procedure was otherwise identical to that of the present study. The ordering of the two sessions was split such that half of the participants in the present study completed this experiment first and half completed the other experiment first.

Results

Comprehension accuracy

Response accuracy was high overall, as can be seen in Figure 1. Cronbach's alpha (a measure of reliability) for the two sets of comprehension questions that appeared with the two different presentation lists was .69 and .78.

Data analysis and descriptive statistics

For the eye-movement data, we used a combination of early and late measures of sentence processing (as defined by Clifton et al., 2007), all for individual words. *First fixation duration* is the amount of time spent the first time a participant looks at a word. It is an early measure that is sensitive to lexical factors such as word frequency and
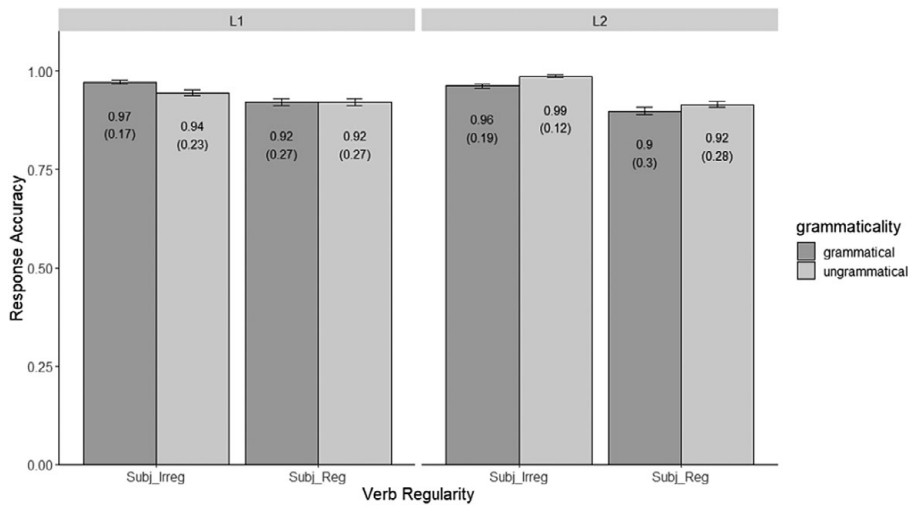
Figure 1. Mean response accuracy for poststimulus comprehension questions (SDs in parenthesis).

polysemy (ibid.). *Total dwell time* is the sum of the durations of all fixations made on the word in question and therefore a late measure. It is often sensitive to higher-level factors related to syntax, semantics, and pragmatics (ibid.). *Regressions to* refers to whether a stimulus word was fixated using an eye regression from an area to the right of the word, whereas *regressions from* refers to whether this area was the starting point for a regressive eye movement back to an area to the left of the word. Regressions are most often viewed as a late measure that reflects reanalysis, but they can reflect earlier processes in some cases, such as when reanalysis is triggered by word-level factors (ibid.). For the present study, there was no previous eye-tracking research to inform our predictions regarding the individual eye-movement measures, but the processing of mood involves morphosyntax and semantics, so it seemed most likely that the effects of interest would appear in the later measures, total dwell time and regressions.

Eye movements were examined for a total of four key words in the stimuli, as illustrated in Figure 2. These included first fixation duration, total dwell time, and regressions to and from the critical verb, which was the embedded clause verb in the subjunctive (grammatical) or indicative mood (ungrammatical), and the two following words, referred to as the critical verb + 1 and critical verb + 2. We also examined the regressions to the subjunctive trigger verb in the main clause.

Descriptive statistics for first fixation duration and total dwell time at the critical verb and the following two words can be found in Figures 3 and 4, and 5 and 6, respectively. Descriptive statistics for regressions to the critical verb, the following

El	ministro	espera	que	los	ciudadanos	aprueben/ *aprueban	su	propuesta.
The	minister	expects	that	the	citizens	approve$_{SUBJ/*IND}$	his/her	proposal.
		TRIGGER VERB				CRITICAL VERB	CRITICAL VERB + 1	CRITICAL VERB + 2

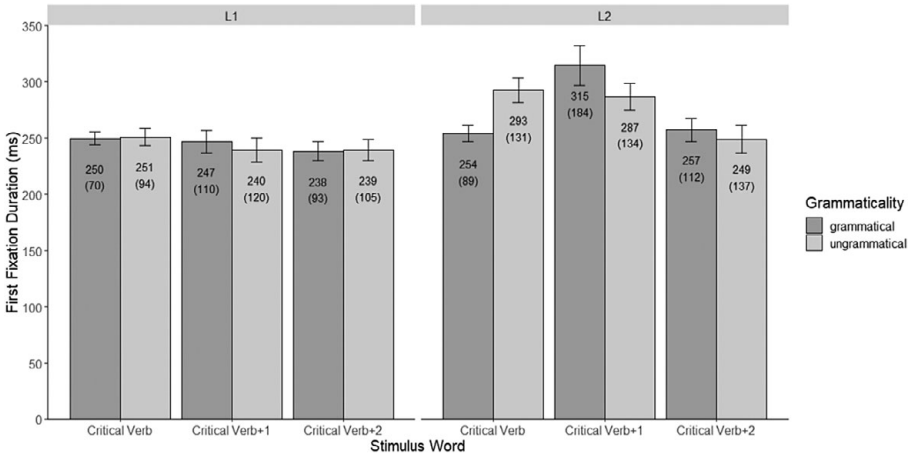**Figure 2.** Stimulus regions of interest.

Figure 3. Mean first fixation duration for irregular verb stimuli (SDs in parenthesis).
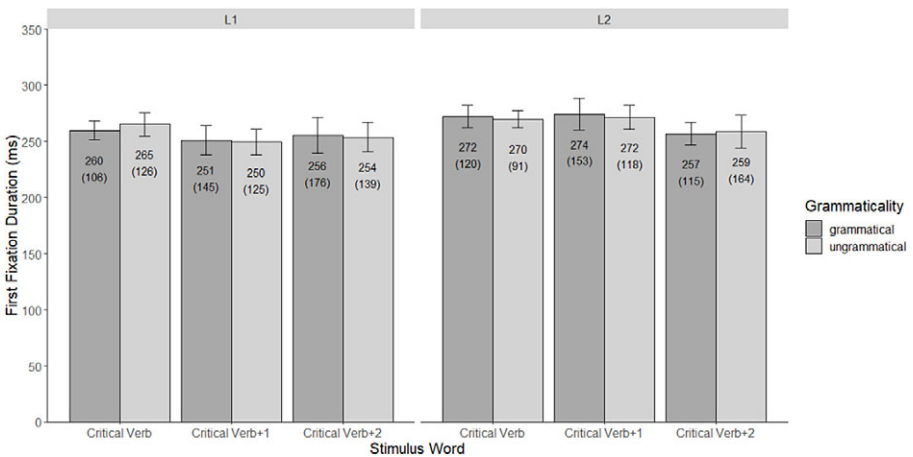

Figure 4. Mean first fixation duration for regular verb stimuli (SDs in parenthesis).

word, and trigger verb can be seen in Figures 7 and 8. Lastly, descriptive statistics for regressions from the critical verb and the following two words can be seen in Figures 9 and 10.

Prior to statistical analysis, we implemented only minimal trimming of the time-based data to remove absolute outliers prior to transformation (Baayen & Milin, 2010). Specifically, fixation values below 100 ms were removed from the first fixation duration and total dwell time data at the recommendation of an anonymous reviewer and because fixations of less than 50 ms appear not to yield useful information (Ihoff & Radach, 1998) and fixations of less than 100 ms are rare (Rayner, 1998) and thus commonly treated as outliers. This affected 2.6% of the total data for first fixation duration and 1.5% of the data for total dwell time. The remaining time-based data was then log-transformed to reduce positive skew.
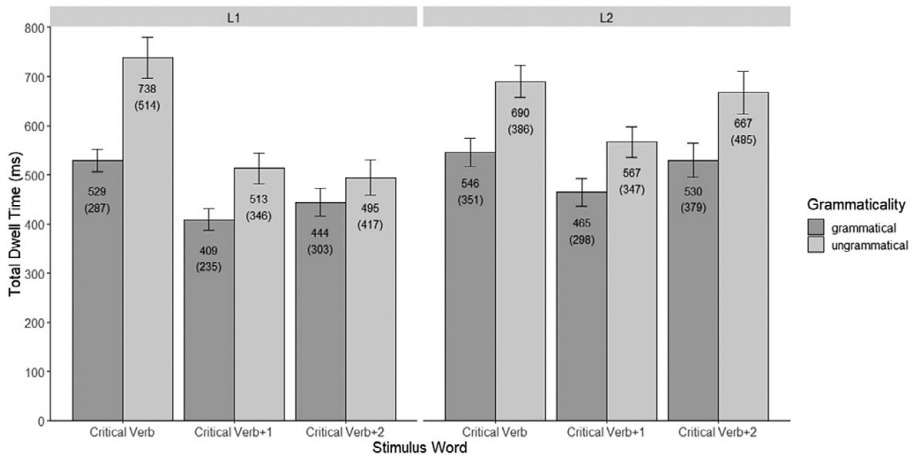
Figure 5. Mean total dwell time for irregular verb stimuli (SDs in parenthesis).
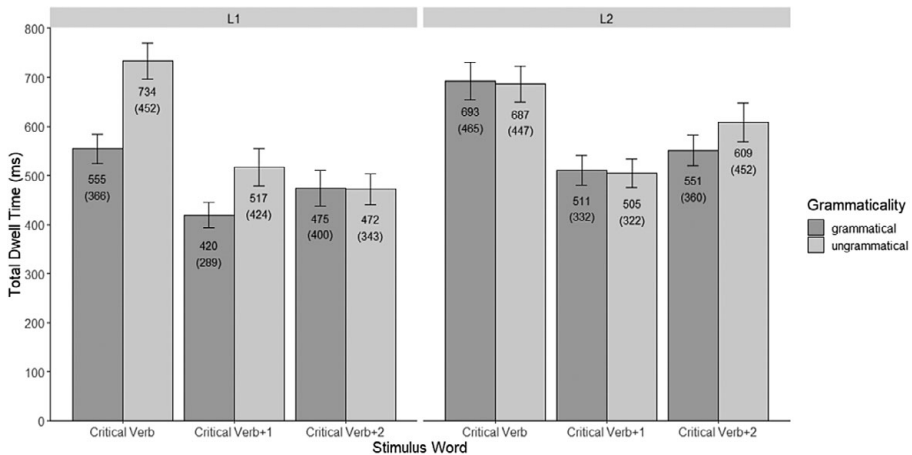

Figure 6. Mean total dwell time for regular verb stimuli (SDs in parenthesis).

First fixation duration and total dwell time were analyzed using linear mixed-effects models using R (R Core Team, 2021) with the *lme4* package (Bates et al., 2015) and keeping the maximal random effect structure whenever possible (following Barr, 2013). *P* values were obtained using Satterthwaite's approximation for degrees of freedom with the *lmerTest* package for R (Kuznetsova et al., 2014). Pairwise comparisons were conducted using the *emmeans* package (Lenth et al., 2018), which employs the Tukey method for multiple comparisons. Alpha was set at .05 for all analyses, *p* = .05 was treated as significant, and interactions with p-values less than .10 were explored as potentially significant, to minimize Type II error likelihood (Larson-Hall, 2010).

Regressive eye movements were analyzed using a Bayesian approach, after an initial attempt to use a mixed-effects logistic regression analysis resulted in most models not converging. This is a common problem with regression data, which are binary and

Figure 7. Regressions to trigger verb, critical verb, and subsequent word with irregular verb stimuli (proportion of trials; SDs in parenthesis).

Figure 8. Regressions to trigger verb, critical verb, and subsequent word with regular verb stimuli (proportion of trials; SDs in parenthesis).

typically contain a high number of zeros, and a Bayesian approach can help with the problem of nonconvergence (Hofmeister & Vasishth, 2014; Husain et al., 2014). Following Kimball et al. (2018), we ran Bayesian models with a maximal random effect structure and no priors using the *brms* package (Bürkner, 2017, 2018). Pairwise comparisons were conducted comparing the probability distribution of the differences that were found to have a true difference in the output of the Bayesian models. If their distribution crossed zero, there was not a true probabilistic difference between conditions or groups.
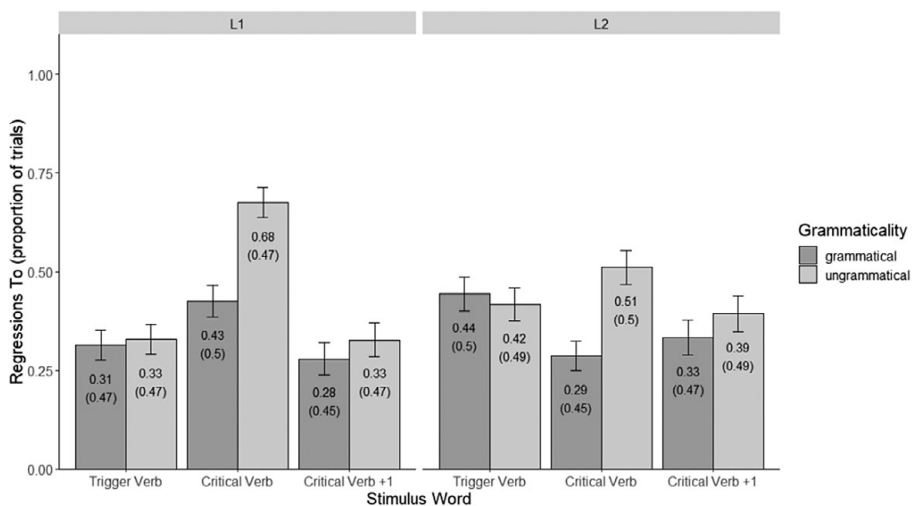
Figure 9. Regressions from the critical verb and subsequent words with irregular verb stimuli (proportion of trials; SDs in parenthesis).

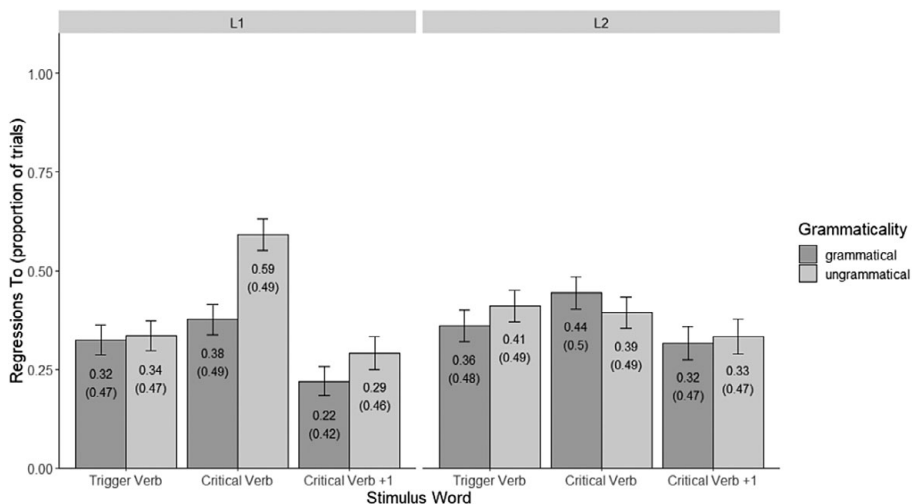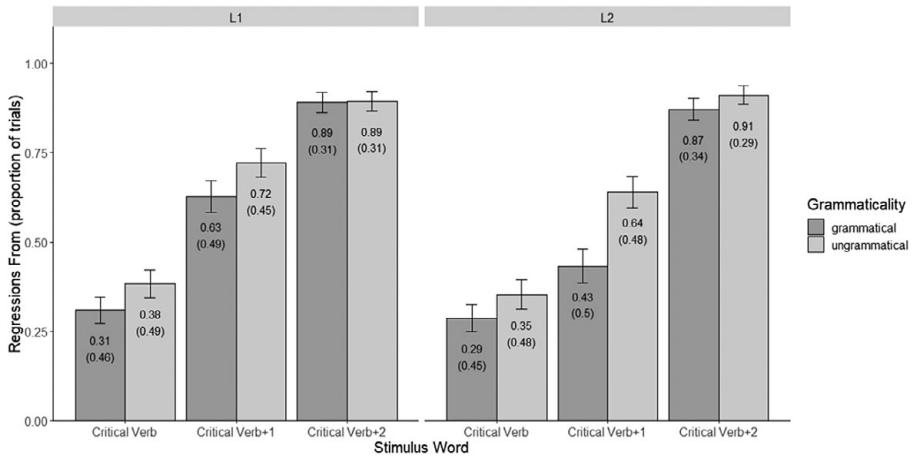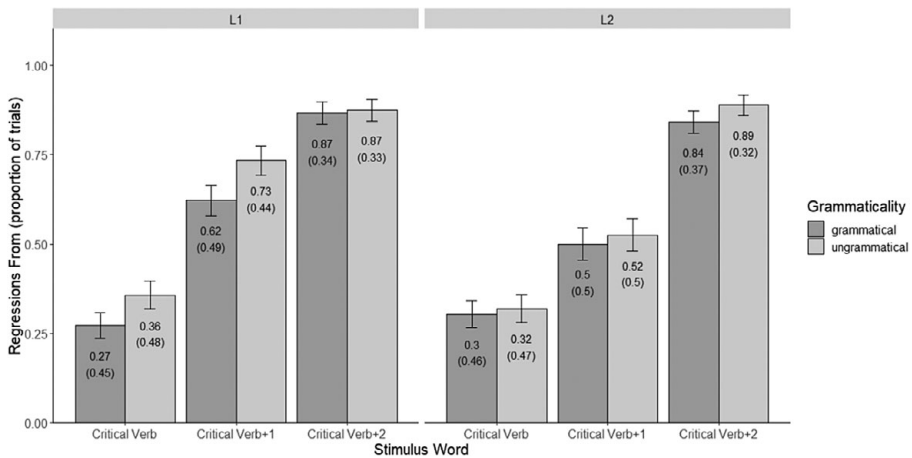**Figure 10.** Regressions from critical verb and subsequent words with regular verb stimuli (proportion of trials; SDs in parenthesis).

Standardized effect sizes were estimated independently from the mixed-effects models, calculated as Cohen's *d* with a correction for dependence between means in the comparisons that were within-subjects (Morris & DeShon, 2002; Wiseheart, 2014).

For all primary analyses, the fixed effects were group (L1, L2), verb regularity (irregular, regular), and grammaticality (grammatical, ungrammatical), and the random effects were subject and item. Verb frequency was included as a covariate in all the models because the irregular verbs were more frequent than the regular verbs, as discussed in the "Materials" section, and will be mentioned only when significant. Experimental session was also included as a covariate in the primary (omnibus) models, but it was never significant, so it is reported in the output tables below but will not be discussed further. The reference levels for the fixed effects were L2, irregular, and grammatical, respectively, although with only two levels of each variable, this

designation did not affect the contrasts examined. The maximal model also included verb regularity, grammaticality, and their interaction as by-subject slopes and grammaticality, group, and their interaction as by-item slopes. When a model did not converge, this random slope structure was simplified incrementally until the model did converge. For more details, see the R code for all primary analyses in the online supplementary materials.

First fixation duration

The output of the statistical analyses of the first fixation duration on the critical verb and two postcritical words can be seen in Table 2. At the critical verb, there was a main effect of grammaticality (effect size of $d = 0.12$), no effect of group ($d = 0.14$), no effect of verb regularity ($d = 0.07$), and a significant interaction of grammaticality with group and of grammaticality with verb regularity. The covariate of frequency showed a predictable effect, with higher frequency verbs showing shorter fixations. Additional analyses were conducted separately by group to explore the interactions. The L1 group

Table 2. First fixation duration for critical verb and two subsequent words: Output from linear mixed-effects models

	Estimate	SE	t	p
Critical verb				
Intercept	5.74	0.08	64.44	0.00
Grammaticality	0.11	0.03	2.96	0.00
Group	0.00	0.05	0.13	0.89
Verb Regularity	0.03	0.03	0.90	0.36
Frequency	−0.03	0.01	−2.93	0.00
Session	−0.08	0.04	−1.99	0.06
Grammaticality × Group	−0.13	0.05	−2.44	0.01
Grammaticality × Verb Regularity	−0.10	0.05	−1.91	0.05
Group × Verb Regularity	−0.03	0.05	−0.62	0.53
Grammaticality × Group × Verb Regularity	0.12	0.07	1.62	0.10
Critical verb +1				
Intercept	5.71	0.11	51.54	0.00
Grammaticality	−0.07	0.05	−1.18	0.24
Group	−0.20	0.08	−2.51	0.01
Verb regularity	−0.14	0.05	−2.59	0.01
Frequency	−0.01	0.01	−0.59	0.55
Session	−0.03	0.04	−0.66	0.51
Grammaticality × Group	0.02	0.08	0.32	0.74
Grammaticality × Verb Regularity	0.07	0.07	0.99	0.32
Group × Verb Regularity	0.15	0.07	1.94	0.05
Grammaticality × Group × Verb Regularity	−0.04	0.11	−0.40	0.68
Critical verb +2				
Intercept	5.48	0.12	45.30	0.00
Grammaticality	−0.05	0.05	−1.05	0.29
Group	−0.05	0.06	−0.81	0.41
Verb Regularity	0.00	0.04	0.08	0.93
Frequency	0.00	0.01	0.03	0.97
Session	−0.01	0.05	−0.33	0.73
Grammaticality × Group	0.04	0.06	0.63	0.52
Grammaticality × Verb Regularity	0.02	0.06	0.31	0.75
Group × Verb Regularity	−0.00	0.06	−0.06	0.94
Grammaticality × Group × Verb Regularity	0.01	0.09	0.12	0.89

showed no effect of grammaticality, estimate = 0.00, SE = 0.02, t = 1.70, p = 0.08, d = 0.04, or verb regularity, estimate = 0.01, SE = 0.02, t = 0.64, p = 0.51, d = 0.13, no interaction, estimate = −0.03, SE = 0.03, t = −1.02, p = 0.30, and the covariate of frequency showed an effect in the expected direction, estimate = −0.03, SE = 0.01, t = −2.94, p = 0.00. The L2 group showed an effect of grammaticality, estimate = 0.11, SE = 0.04, t = 2.84, p = 0.00, d = 0.19, no effect of verb regularity, estimate = 0.03, SE = 0.04, t = 0.84, p = 0.39, d = 0.03, a borderline significant interaction, estimate = −0.10, SE = 0.05, t = −1.84, p = 0.06, and the covariate of frequency showed an effect in the expected direction, estimate = −0.03, SE = 0.01, t = −2.08, p = 0.03. Pairwise comparisons conducted to explore the potential interaction confirmed a significant effect of grammaticality with the irregular verb stimuli, estimate = 0.11, SE = 0.04, t = 2.50, p = 0.02, d = .37, that was not present with regular verb stimuli, estimate = 0.01, SE = 0.04, t = 0.32, p = 0.74, d = .02.

At the critical verb + 1, there were main effects of verb regularity (d = 0.07) and group (d = 0.28), as well as a significant interaction of verb regularity with group, but no effect of grammaticality (d = 0.07). Pairwise comparisons revealed a significant main effect of verb regularity for the L2 group, estimate = −0.12, SE = 0.05, t = −2.14, p = 0.01, d = .20, that was not present with the L1 group, estimate = 0.00, SE = 0.05, t = −0.15, p = 0.88, d = .05. The L2 group showed generally longer first fixation durations for irregular verbs than regular ones, but it is difficult to interpret the effect because the irregular and regular verb stimuli were different items not designed to be directly compared to each other, but rather in terms of an interaction with grammaticality (for which the carefully controlled stimulus conditions were identical except for the grammaticality of the verb).

Finally, at the critical verb + 2, which was also the last word of the sentence, there were no significant effects (group: d = 0.07; grammaticality: d = 0.02; verb regularity: d = 0.08) or interactions. Neither group showed any lingering effect of mood at this point in the stimulus.

Total dwell time

The output of the statistical analyses of the total dwell time on the critical and two postcritical words can be seen in Table 3. At the critical verb, there was a main effect of grammaticality (d = 0.39), a main effect of verb regularity (d = 0.19), no effect of group (d = 0.04), a significant interaction of grammaticality with verb regularity, a significant interaction of verb regularity with group, and a significant three-way interaction. The covariate of frequency showed the expected effect, with higher frequency verbs showing shorter fixations. Additional analyses were conducted separately by group to explore the interactions. The L1 group showed an effect of grammaticality, estimate = 0.23, SE = 0.10, t = 2.30, p = 0.03, d = 0.64, but no effect of verb regularity, estimate = −0.03, SE = 0.05, t = −0.59, p = 0.55, d = 0.06, and no interaction, estimate = 0.03, SE = 0.11, t = 0.34, p = 0.73. The covariate of frequency showed an effect in the expected direction, estimate = −0.10, SE = 0.02, t = −3.84, p = 0.00. The L2 group showed a main effect of grammaticality, estimate = 0.24, SE = 0.06, t = 3.92, p = 0.00, d = 0.23, no effect of verb regularity, estimate = 0.14, SE = 0.07, t = 1.89, p = 0.06, d = 0.35, and a significant interaction of grammaticality with verb regularity, estimate = −0.24, SE = 0.08, t = −2.79, p = 0.00. The covariate of frequency showed an effect in the expected direction, estimate = −0.16, SE = 0.04, t = −3.97, p = 0.00. Pairwise comparisons revealed that the L2 group was sensitive to grammatical mood with irregular verb

Table 3. Total dwell time for critical verb and two subsequent words: Output from linear mixed-effects models

	Estimate	SE	t	p
Critical verb				
Intercept	6.63	0.20	32.28	0.00
Grammaticality	0.24	0.07	3.49	0.00
Group	0.04	0.10	0.39	0.69
Verb Regularity	0.16	0.06	2.40	0.01
Frequency	−0.13	0.02	−4.74	0.00
Session	−0.01	0.10	−0.15	0.88
Grammaticality × Group	−0.01	0.09	−0.10	0.91
Grammaticality × Verb Regularity	−0.24	0.08	−2.91	0.00
Group × Verb Regularity	−0.21	0.08	−2.61	0.00
Grammaticality × Group × Verb Regularity	0.27	0.11	2.38	0.01
Critical verb +1				
Intercept	5.98	0.21	27.45	0.00
Grammaticality	0.17	0.07	2.29	0.02
Group	−0.09	0.09	−1.01	0.31
Verb Regularity	0.04	0.08	0.54	0.58
Frequency	−0.03	0.03	−0.84	0.40
Session	0.04	0.09	0.47	0.63
Grammaticality × Group	−0.01	0.10	−0.10	0.91
Grammaticality × Verb Regularity	−0.18	0.10	−1.79	0.07
Group × Verb Regularity	−0.05	0.10	−0.58	0.56
Grammaticality × Group × Verb Regularity	0.19	0.14	1.36	0.17
Critical verb +2				
Intercept	6.04	0.26	22.54	0.00
Grammaticality	0.16	0.09	1.67	0.10
Group	−0.14	0.14	−1.02	0.31
Verb Regularity	0.07	0.09	0.81	0.41
Frequency	0.00	0.03	0.17	0.85
Session	−0.06	0.12	−0.48	0.62
Grammaticality × Group	−0.08	0.13	−0.59	0.55
Grammaticality × Verb Regularity	−0.11	0.10	−1.03	0.30
Group × Verb Regularity	−0.03	0.10	−0.32	0.74
Grammaticality × Group × Verb Regularity	0.04	0.15	0.32	0.74

stimuli, estimate = 0.24, $SE = 0.06$, $t = 3.96$, $p = 0.00$, $d = 0.55$, but not with regular verb stimuli, estimate = 0.00, $SE = 0.07$, $t = 0.02$, $p = 0.97$, $d = 0.01$, and the covariate of frequency was predictably significant in both models.

At the critical verb + 1, there was a main effect of grammaticality ($d = .23$), there was no effect of group ($d = 0.15$) or verb regularity ($d = 0.01$), and the interaction of grammaticality with verb regularity approached significance. Follow-up analyses of each verb type were run separately to explore the potential interaction. These revealed a grammaticality effect with the irregular verb stimuli, estimate = 0.17, $SE = 0.07$, $t = 2.48$, $p = 0.01$, $d = 0.34$, that was not present with the regular verb stimuli, estimate = −0.00, $SE = 0.07$, $t = −0.00$, $p = 0.07$, $d = 0.12$. Thus, online sensitivity to mood carried over to the post-critical word regardless of group, but only with the irregular verb stimuli.

Finally, at the critical verb + 2 region, there were no significant effects (group: $d = 0.30$; grammaticality: $d = 0.22$; verb regularity: $d = 0.02$) or interactions. Neither group showed any lingering sensitivity to mood with either type of verb.

Regressions to

As previously stated, a Bayesian approach was adopted for analysis of the regression data. In frequentist approaches that rely on median and average values, confidence intervals are based only on the data as it is. However, in a Bayesian approach, a credible interval (CrI) incorporates prior probability distributions to signal an interval within which an unobserved parameter value falls. True differences, which can be said to be equivalent to statistically significant differences in frequentist approaches, are signaled with the upper and lower 95% credible intervals values not crossing zero.

The output of the statistical analyses of the regressions to the trigger verb, critical verb, and subsequent word can be seen in Table 4. Regressions to the critical verb + 2 word were not possible because it was the last word of the sentence. At the trigger verb there were no true differences of group, grammaticality, verb regularity, frequency, or session (group: $d = 0.17$; grammaticality: $d = 0.03$; verb regularity: $d = 0.03$), nor were there any interactions.

Table 4. Regressions to the trigger verb, critical verb and critical verb + 1: Output from Bayesian logistic mixed-effects models

	Estimate	SE	l-95% CrI	u-95% CrI	Rhat
Trigger Verb					
Intercept	0.38	0.71	−1.01	1.77	1.00
Grammaticality	−0.10	0.27	−0.63	0.44	1.00
Group	−0.42	0.35	−1.13	0.26	1.00
Verb Regularity	−0.28	0.30	−0.86	0.31	1.00
Frequency	−0.17	0.12	−0.41	0.07	1.00
Session	−0.17	0.30	−0.76	0.43	1.00
Grammaticality × Group	0.21	0.39	−0.55	0.96	1.00
Grammaticality × Verb Regularity	0.34	0.36	−0.36	1.05	1.00
Group × Verb Regularity	0.35	0.38	−0.39	1.09	1.00
Grammaticality × Group × Verb Regularity	−0.47	0.52	−1.49	0.55	1.00
Critical verb					
Intercept	−0.63	0.57	−1.75	0.50	1.00
Grammaticality	0.86	0.33	0.22	1.53	1.00
Group	0.75	0.32	0.11	1.37	1.00
Verb Regularity	0.68	0.27	0.16	1.22	1.00
Frequency	−0.07	0.10	−0.26	0.11	1.00
Session	−0.11	0.25	−0.60	0.39	1.00
Grammaticality × Group	0.36	0.48	−0.54	1.30	1.00
Grammaticality × Verb Regularity	−1.16	0.40	−1.95	−0.38	1.00
Group × Verb Regularity	−0.98	0.37	−1.69	−0.26	1.00
Grammaticality × Group ×x Verb Regularity	0.92	0.55	−0.17	2.00	1.00
Critical verb + 1					
Intercept	−1.46	0.83	−3.10	−0.15	1.00
Grammaticality	0.54	0.29	−3.10	0.15	1.00
Group	−0.18	0.43	−1.05	0.69	1.00
Verb Regularity	0.19	0.32	−0.44	0.84	1.00
Frequency	0.05	0.14	−0.24	0.34	1.00
Session	−0.11	0.36	−0.82	0.58	1.00
Grammaticality × Group	−0.38	0.44	−1.23	0.46	1.00
Grammaticality × Verb Regularity	−0.57	0.42	−1.39	0.25	1.00
Group × Verb Regularity	−0.42	0.46	−1.33	0.47	1.00
Grammaticality × Group × Verb Regularity	0.73	0.62	−0.46	1.96	1.00

Note: True differences are signaled by the lower and upper credible intervals (CrIs) not crossing zero, i.e., both values being positive or negative.

At the critical verb, there were main effects of grammaticality ($d = 0.25$), group ($d = 0.22$), and verb regularity ($d = 0.06$), and interactions of verb regularity with group and with grammaticality. Additional analyses were conducted separately for each group to explore the interactions. The L1 group showed an effect of grammaticality, estimate = 1.23, $SE = 0.37$, 95% CrI [0.52, 1.96], $d = 0.34$, but no effect of verb regularity, estimate = −0.29, $SE = 0.26$, 95% CrI [−0.81, 0.22], $d = 0.15$, and no interaction, estimate = −0.23, $SE = 0.39$, 95% CrI [−1.01, 0.54]. The L2 group also showed a main effect of grammaticality, estimate = 0.90, $SE = 0.32$, 95% CrI [0.26, 1.55], $d = 0.13$, plus they also showed an effect of verb regularity, estimate = 0.70, $SE = 0.28$, 95% CrI [0.15, 1.25], $d = 0.06$, and an interaction, estimate = −1.15, $SE = 0.43$, 95% CrI [−1.99, −0.33]. A follow-up analysis run with verb regularity separately with the L2 data revealed an effect of grammaticality with the irregular verb stimuli, estimate = 0.87, $SE = 0.36$, 95% CrI [0.16, 1.53], $d = 0.40$, but not with the regular verb stimuli, estimate = −0.27, $SE = 0.36$, 95% CrI [−1.00, 0.44], $d = 0.06$.

At the critical verb + 1, there were no true differences of group, grammaticality, or verb regularity (group: $d = 0.14$; grammaticality: $d = 0.12$; verb regularity: $d = 0.10$), nor were there any interactions.

Regressions from

The output of the statistical analyses of the regressions from the critical verb and the two following words can be seen in Table 5. At the critical verb, there were no main effects (group: $d = 0.03$; grammaticality: $d = 0.12$; verb regularity: $d = 0.05$) or interactions.

At the critical verb + 1, there were effects of grammaticality ($d = 0.24$) and group ($d = 0.31$), but no effect of verb regularity ($d = 0.03$). Moreover, there was an interaction of verb regularity with grammaticality, so additional analyses were conducted for regular and irregular verb stimuli separately. The irregular verb stimuli showed an effect of grammaticality, estimate = 0.97, $SE = 0.27$, 95% CrI [0.46, 1.50], $d = 0.27$, but the regular verb stimuli did not, estimate = 0.03, $SE = 0.29$, 95% CrI [-0.52, 0.59], $d = 0.10$.

At the critical verb + 2, there were no main effects of group, grammaticality, or verb regularity (group: $d = 0.01$; grammaticality: $d = 0.07$; verb regularity: $d = 0.08$), nor were there any interactions.

Results summary

The main results of this experiment were as follows:

- Only the L2 group showed online sensitivity to mood in the early measure of first fixation duration. This was at the critical verb and occurred only with irregular verbs.
- Both groups showed sensitivity to mood in the later measure of total dwell time with the irregular verb stimuli. This was at both the critical verb and the following word. In addition, only the L1 group showed sensitivity to mood with the regular verb stimuli, and this was only on the critical verb, with no spillover.
- Similarly, both groups showed sensitivity to mood in the regressions to the critical verb with the irregular verb stimuli and only the L1 group showed the effect with the regular verb stimuli.
- Both groups showed sensitivity to mood in the regressions from the postcritical word with the irregular verb stimuli, and neither group showed the effect with the regular verb stimuli.

Table 5. Regressions from critical verb and two subsequent words: Output from Bayesian logistic mixed-effects models

	Estimate	SE	l-95% CrI	u-95% CrI	Rhat
Critical verb					
Intercept	−1.02	0.58	−2.15	0.14	1.00
Grammaticality	0.18	0.29	−0.40	0.76	1.00
Group	0.20	0.34	−0.46	0.87	1.00
Verb Regularity	0.07	0.29	−0.52	0.66	1.00
Frequency	−0.08	0.10	−0.27	0.11	1.00
Session	0.14	0.25	−0.35	0.63	1.00
Grammaticality × Group	0.16	0.40	−0.63	0.95	1.00
Grammaticality × Verb Regularity	−0.19	0.39	−0.98	0.58	1.00
Group × Verb Regularity	−0.34	0.39	−1.12	0.43	1.00
Grammaticality × Group × Verb Regularity	0.26	0.54	−0.82	1.35	1.00
Critical verb + 1					
Intercept	−0.50	0.68	−1.84	0.85	1.00
Grammaticality	0.97	0.26	0.46	1.49	1.00
Group	0.83	0.33	0.20	1.49	1.00
Verb Regularity	0.44	0.30	−0.16	1.04	1.00
Frequency	0.00	0.12	−0.23	0.24	1.00
Session	−0.28	0.29	−0.85	0.27	1.00
Grammaticality × Group	−0.49	0.37	−1.25	0.20	1.00
Grammaticality × Verb Regularity	−0.94	0.37	−1.68	−0.24	1.00
Group × Verb Regularity	−0.27	0.37	−0.98	0.46	1.00
Grammaticality × Group × Verb Regularity	0.85	0.52	−0.15	1.89	1.00
Critical verb + 2					
Intercept	1.46	1.31	−1.10	4.03	1.0
Grammaticality	0.26	0.32	−0.38	0.88	1.0
Group	−0.06	0.60	−1.21	1.16	1.0
Verb Regularity	0.26	0.42	−0.58	1.08	1.0
Frequency	0.23	0.21	−0.18	0.65	1.0
Session	−0.69	0.59	−1.84	0.42	1.00
Grammaticality × Group	0.34	0.44	−0.52	1.21	1.0
Grammaticality × Verb Regularity	−0.11	0.46	−1.00	0.80	1.0
Group × Verb Regularity	−0.26	0.44	−1.13	0.59	1.0
Grammaticality × Group × Verb Regularity	−0.40	0.63	−1.64	0.84	1.0

Note: True differences are signaled by the lower and upper credible intervals (CrIs) not crossing zero, i.e., both values being positive or negative.

- The covariate of verb frequency showed a predictable effect with the time-based measures of first fixation duration and total dwell time, in which fixation times on the critical verb were shorter when the verb was of higher frequency. This suggests that frequency did account for some of the variance in those models and including it as a covariate potentially helped to clarify some of the results. Frequency did not appear to make a difference in the models for the regression data.

Discussion

The first research question for the present study asked if advanced L2 speakers of Spanish were sensitive to grammatical mood during online sentence comprehension. Our findings suggest that, in the most basic sense, the answer to this question is affirmative. The L2 participant group in this study showed the effect of interest in five analyses of four different eye-movement measures: first fixations were longer on the critical verb, total dwell times were longer on the critical verb and the following word,

there were more trials with regressions to the critical verb, and there were more trials with regressions from the postcritical word, all with ungrammatical stimuli versus grammatical stimuli. Of course, the L2 group only showed these effects with the irregular verb stimuli, a point that will be discussed further in the following text, in the discussion of the second research question. Nevertheless, the data from this study show that it is possible for L2 learners to integrate verbal mood morphology during online sentence comprehension. To our knowledge, this is the first empirical study to provide such evidence.

The first research question also proposed a comparison of L2 and L1 participant groups with regard to online sensitivity to Spanish mood. Both groups showed the expected effect across multiple eye-movement measures and across two stimulus words, in the case of total dwell time, so they were similar overall. However, one point of difference was that only the L2 group showed early sensitivity to mood in the first fixation duration measure. Early sensitivity to mood morphology was not predicted even with the L1 group, given how many different linguistic factors are involved in the processing of mood in Spanish, so the fact that it was observed here among L2 learners suggests that nonnative processing of mood morphology was very efficient. Hence, we interpret the results of this study as evidence of nativelike sensitivity to mood in L2 sentence processing. At the same time, there was evidence of L1/L2 differences in processing at the word level, to be discussed next.

The second research question for the present study asked if the regularity of verbs with mood morphology affected online sensitivity to the form among advanced L2 users. Our findings suggest that verb regularity was indeed very important in this experiment. The L2 participant group in this study showed robust online sensitivity to verbal mood morphology, as discussed previously under the first research question, but only with irregular verb stimuli. There was no evidence of online sensitivity to mood morphology with regular verbs.

The second research question also proposed a comparison of L2 and L1 participant groups with regard to the role of verb regularity in the online processing of verbal mood morphology. Here there was a notable difference between the two groups in terms of the degree to which verb regularity affected their processing of mood. With the L2 group, the effect of verb regularity was absolute, meaning there was no evidence of online sensitivity to mood with regular verb stimuli and robust evidence of online sensitivity to mood with irregular verbs. The L1 group, however, showed robust online sensitivity to mood with both verb types. Nevertheless, this sensitivity was slightly more robust with the irregular verb stimuli than with the regular verb stimuli. In total dwell times, the effect was present across two words with the irregular verb stimuli, but only on the critical word with the regular verb stimuli. And with the regressions from the post-critical word, the effect was only observed with the irregular verb stimuli. Thus, it appears that verb regularity can affect the L1 processing of mood as well, albeit in a more subtle manner than with L2 processing. To our knowledge, this is the first study to observe an apparent role of verb regularity in the native processing of Spanish mood.

The outcome of the present study differs from that of Cameron (2011, 2017), who conducted the only previous study of the L2 acquisition of the Spanish subjunctive using a real-time method (self-paced reading) and found no evidence of online sensitivity to verbal mood among L2 users of Spanish. One very plausible explanation for the difference is that the present study included stimuli with both regular and irregular verbs, whereas the Cameron study used only regular verbs. The results of the two studies could be seen as consistent in this regard, as both found a lack of online sensitivity to mood with regular verb stimuli among L2 users, even though L1 users

showed the predicted effect. Although this one difference could explain the apparently different findings of the two studies in and of itself, there are at least two other notable methodological differences that might have also been important. First, the "trigger" verbs used in the present study (e.g., *querer* "to want," *esperar* "to hope or expect") are more frequently associated with the subjunctive mood than the expressions of certainty used in the Cameron study (e.g., *Es possible que* "It is possible that," *Es probable que* "It is probable that"; subjunctive frequency data from Davies, 2006, p. 142). The native speakers showed the expected online effects in both studies, however, so the linguistic context does not appear to have been a limiting factor in any absolute sense. Second, the participants in the present study were likely of higher proficiency than even the most advanced group in the Cameron study (mean score 45/50 vs. 36/50 on proficiency tests that were both based on the DELE). Further research is needed to investigate the role of the type of subjunctive mood trigger in the stimulus sentences and to explore the role of L2 proficiency, while keeping in mind the important role of verb form regularity.

Turning now to the theoretical implications of the outcome of the present study, we found no evidence of deficiencies in the nonnative processing of sentence-level syntax and morphosyntax of the type proposed by the Shallow Structure Hypothesis (Clahsen & Felser 2006, 2018). Although the theory has not dealt with grammatical mood specifically, verb morphology and the abstract grammatical details it conveys are of key interest and mood thus seems to fit. As outlined in the background of this paper, the processing of grammatical mood involves verb morphology, sentence-level semantics, and a nonlocal dependency (with three words of separation) between the lexical-semantics of the verb in the matrix clause and mood morphology on the verb in an embedded clause. Despite this linguistic complexity and the need to integrate multiple sources of abstract grammatical information across clauses, the advanced L2 users in the present study showed nativelike online sensitivity to grammatical mood with irregular verbs. Thus, L2 sentence processing was not intrinsically or inescapably shallow, as this would entail a lack of sensitivity in all contexts, regardless of verb regularity.

Nevertheless, L2 processing did appear to be limited to some degree, as there was no evidence of online sensitivity to grammatical mood with regular verbs. This stood in contrast with the results from the L1 group, which showed the expected effects, although these were not quite as robust as with irregular verbs (present in four analyses with irregular verbs vs. two measures with regular verbs). The clear pattern of nativelike processing of mood with irregular verbs and no evidence of online processing of mood with regular verbs that was observed among the L2 group is consistent with a body of previous research on the adult L2 acquisition of Spanish mood that has also identified verb form regularity as an important factor (Collentine, 1997; Gallego & Pozzi, 2018; Gudmestad, 2006, 2012a; Lubbers-Quesada, 1998). The prior work was conducted using a variety of offline methods as opposed to a real-time measure like eye tracking and the differences between regular and irregular verbs were often more graded than in the present study, but the results were generally similar in that more nativelike behavior was seen with irregular verbs than with regular ones. The outcome of the present study was also broadly consistent with the results of one prior study that had also examined the role of verb regularity in nonnative sentence processing: Pliatsikas and Marinis (2013) found that native speakers and high-proficiency L2 learners were slower to process the English past tense with regular verbs than with irregular verbs. However, the L2 group still showed some degree of online sensitivity to the past tense with regular verbs, unlike in the present study.

In considering possible explanations for the role of verb regularity, three word-level factors that might be of importance and that are not mutually exclusive are word frequency, perceptual salience, and morphological regularity, each of which will be considered here in turn. The first factor is word frequency: As is typical, the irregular verbs used in the present study were more frequent than the regular ones and this could potentially lead to differences in processing speed at the word level that might translate into difficulty in processing morphosyntax at the sentence level (Hopp, 2016; Jegerski & Fernandez Cuenca, 2019). For this reason, frequency of the critical verbs was included as a covariate in our analyses. It is therefore not likely that verb frequency played a role in the different outcomes for regular and irregular verb stimuli in the present study, so form salience and form regularity (both to be discussed in the following text) were probably more important. Speaking more broadly, however, it is still possible that the higher frequency of irregular verbs compared to regular ones is a factor in the acquisition of Spanish mood (Pinker, 1999), it is just that it does not appear to have been a factor in this particular experiment, probably because both sets of verbs were of high frequency (cf. Hopp, 2016; Jegerski & Fernández Cuenca, 2019; Jiang & Botana, 2009) and not very different from each other in relative terms, even though the difference was statistically significant.

A second factor is perceptual salience. As has been pointed out by a number of other researchers working on the acquisition of the Spanish subjunctive (Collentine, 1997; Gallego & Pozzi, 2018; Gudmestad, 2006, 2012a, Lubbers-Quesada, 1998), Spanish mood is more visually and acoustically salient with irregular verbs than with regular ones. The difference between indicative and subjunctive forms is greater in terms of the number of written letters that vary (e.g., *tiene*/*tenga* "s/he has[IND/SUJB]" vs. *habla*/*hable* "s/he speaks[IND/SUJB]") because both regular and irregular verbs have a thematic vowel shift to indicate mood (from *e*/*i* to *a* or vice versa), but irregular verbs also have a stem change, so it could be that irregular forms are simply more likely to be seen and processed. Regarding acoustic salience, the single vowel that most often distinguishes regular indicative and subjunctive verbs is the /a/ – /e/ contrast, which occurs in an unstressed, final syllable in spoken Spanish. In English, most vowels would be neutralized to schwas in that phonetic environment, so English-dominant bilinguals may tend toward neutralization of such vowels in Spanish as well (e.g., Colantoni et al., 2020). As verb regularity and salience are inevitably linked in Spanish due to irregularity affecting verb stems, salience potentially has a role in any observed effect of verb regularity, including in the results of the present study. And when it comes to real-time processing by nonnative readers, as with the experimental task for the present study, form salience may be of particular importance (Hopp & León Arriaga, 2016; Jegerski, 2015). Salience might also explain the results of the L1 group in the present study, who showed more robust online sensitivity to mood with irregular verbs than with regular ones. Hence, we think that salience is one important reason why verb regularity affected the processing of mood in our study, to the extent that there was no evidence of online sensitivity to mood with regular verbs with L2 users.

A third potential explanation for the observed difference in the online L2 processing of mood with regular versus irregular verbs is form regularity. Dual mechanism theories of morphological processing propose that regular inflected forms are accessed using a morphological rule that assembles the component morphemes, while irregular forms are accessed as whole word forms with separate entries in the lexicon (Pinker, 1999). Later versions of the Shallow Structure Hypothesis (Clahsen & Felser, 2018, p. 2) have proposed a secondary claim that L2 learners fail to use morphological rules to

decompose complex words the way that L1 users do, in addition to the primary claim of the theory regarding sentence processing. Under such an account, both irregular and regular forms would be stored equally as whole-word entries in the lexicon, which does not explain why only irregular verbs were associated with online sensitivity to mood in the present study. The more moderate version of this claim seems to fit with our results, as it proposes L1/L2 difference as more a matter of degrees rather than an absolute difference: L2 learners may apply the same rules as L1 users, but they do so more slowly (Kirkici & Clahsen, 2013). Under this account, the L2 users in the present study might have initiated rule-based processing of Spanish mood with regular verbs, but they were not as efficient as the L1 users in doing so. This subtle difficulty may have compounded with form salience to lead to more pronounced effects on eye movements. However, it is important to note that the L2 group in this study showed no evidence of online sensitivity to mood with regular verbs at all, so this is entirely speculative and the data do not exactly fit with the theoretical claim in question. Even if this is interpreted as one point of potential compatibility, the Shallow Structure Hypothesis as a whole cannot account for our results, as its "core claim" (Clahsen & Felser, 2018, p. 1) pertains to difficulty at the level of phrases and sentences rather than within words. Its authors have also specified that morphology in individual words should present less difficulty than syntax and morphosyntax in sentences (Clahsen & Felser, 2006, p. 35), which is the opposite of the pattern observed in the present study.

On the whole, the outcome of the present investigation suggests that the primary processing difficulty for the L2 participants in this study originated at the word level, not at the sentence level. As long as an irregular verb form facilitated the processing of mood at the word level (because of greater form salience and possibly because of more efficient processing of irregular forms), subsequent sentence level processing of mood proceeded in a nativelike fashion. Our results therefore seem to go against a theoretical account of L2 processing that emphasizes difficulty with sentence-level morphosyntax (i.e., Clahsen & Felser, 2018), as this would predict difficulty processing verbal mood in all contexts, regardless of verb regularity. Rather, the results of the present study are more consistent with the Lexical Bottleneck Hypothesis (Hopp, 2014, 2018), which proposes that difficulty at the word level can often be the critical limitation and that apparent difficulty with sentence processing is an indirect effect of word level difficulty.

An unexpected result of this experiment was that the L2 group showed early sensitivity to mood in the first fixation duration measure, but the L1 group did not. Mood effects were predicted to show up primarily in later measures like total dwell time and regressions because the effects arise from a semantic and morphosyntactic dependency across clause boundaries. It is also generally expected for L1 processing to be more efficient than L2 processing, although there is evidence that L2 learners sometimes read faster than their L1 counterparts (e.g., Felser et al., 2003; Kaan et al., 2015). One possible reason for the difference between the groups might be that the L2 group was more attuned to the subjunctive mood while reading in Spanish because the form receives so much attention in world language classrooms. Alternatively, an anonymous reviewer suggested that the L1 group might have experienced L1 attrition due to immersion in their L2, English, at the time of testing. It is true that other research has shown that L1 attrition can occur in similar L2 immersion contexts with at least some aspects of sentence processing (Chamorro et al., 2016; Dussias & Sagarra, 2007), but we think this explanation is less plausible in the case of the present study because the L2 group was also immersed in English at the time of testing, so the two groups would have had similar results.

Finally, one important limitation of this study that potentially affects the general-izability of the findings is that the stimuli represented only a narrow range of the many different uses of Spanish mood. Only 13 different irregular verbs were included because of the need to control word length across indicative and subjunctive forms, and 6 of the 13 verbs were of the –*tener* type (e.g., *tener* "to have," *mantener* "to maintain"), which all follow the same pattern of inflection for mood, despite being considered irregular. In addition, the trigger verbs appearing in the main clause of the stimuli were all strong triggers, showing only minimal variability, but subjunctive use is known to vary in many other contexts. Further research is therefore needed to determine if the present findings might be replicated widely or may be limited to a relatively small set of verbs and contexts. A second limitation is that irregular verbs, the type associated with L1-like processing in this study, are of limited number in the Spanish lexicon (although they do tend to be of much higher frequency than regular verbs). This means that L2 processing of the subjunctive mood, although in theory possible, occurs in only a minority of cases in the real world. We therefore suggest that one practical implication of the present study—which contributes to a growing body of evidence that verb form regularity can play a role in the acquisition of verb morphosyntax—is that future research should examine the potential application of the observed role of morphological regularity in the context of language instruction. For example, in an input-based teaching method like Processing Instruction (VanPatten & Cadierno, 1993), irregular verb forms might increase the likelihood that learners will process mood in the input, thereby improving the potential for acquisition.

In conclusion, the main finding of this empirical study was that advanced profi-ciency L2 readers showed nativelike processing of Spanish mood with irregular verb stimuli and no evidence of online sensitivity to mood with regular verb stimuli. A comparison group of L1 readers showed the expected effect with both types of stimuli, but the effect was slightly more robust with irregular verbs than with regular verbs. Hence, it appears that form regularity played an important role, particularly in L2 processing. We have argued that this was due to the greater visual salience of Spanish subjunctive forms with irregular verbs versus regular verbs and that difficulty process-ing rule-based regular verbal morphology may have played a role as well. In any case, the results suggest that L2 processing difficulty originated with word-level factors, consistent with the Lexical Bottleneck Hypothesis (Hopp, 2014, 2018), and that sentence-level processing has the potential to be nativelike, which appears to go against claims of generalized difficulty in syntactic and morphosyntactic processing (Clahsen & Felser, 2006, 2018).

Supplementary Materials. To view supplementary material for this article, please visit http://doi.org/10.1017/S027226312200016X.

References

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*, 12–28.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Blake, R. (1983). Mood selection among Spanish-speaking children, ages 4 to 12. *The Bilingual Review*, *10*, 21–32.

Bosque, I. (2012). Mood: Indicative vs. subjunctive. In J. I. Hualde, A. Olarrea, & E. O'Rourke (Eds.), *The Handbook of Hispanic Linguistics* (pp. 373–394). Wiley-Blackwell.

Borgonovo, C., Bruhn de Garavito, J. & Prévost, P. (2005). Acquisition of mood distinctions in L2 Spanish. In A. Brugos, M. R. Clark-Cotton, & S. Ha (Eds.), *Proceedings of the 29th annual Boston University Conference on Language Development* (pp. 97–108). Cascadilla Press.

Borgonovo, C., & Prévost, P. (2003). Knowledge of polarity subjunctive in L2 Spanish. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Boston University Conference on Language Development* (Vol. *1*, pp. 150–161). Cascadilla Press.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*, 45–50.

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Bürkner, P. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*, 395–411.

Cameron, R. (2011). Native and nonnative processing of modality and mood in Spanish [Unpublished doctoral dissertation]. Florida State University, Tallahassee.

Cameron, R. (2017). Lexical preference and online processing of Spanish subjunctive. *Linguistics Unlimited*, *2*, 1–24.

Chamorro, G., Sorace, A., & Sturt, P. (2016). What is the source of L1 attrition? The effect of recent L1 re-exposure on Spanish speakers under L1 attrition. *Bilingualism: Language and Cognition*, *19*, 520–532.

Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*, 107–126.

Clahsen, H., & Felser, C. (2018). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, *40*, 1–14.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–372). Elsevier.

Colantoni, L., Martínez, R., Mazzaro, N., Pérez-Leroux, A. T., & Rinaldi, N. (2020). A phonetic account of Spanish-English bilinguals' divergence with agreement. *Languages*, *5*, 58.

Collentine, J. (1995). The development of complex syntax and mood-selection abilities by intermediate-level learners of Spanish. *Hispania*, *78*, 123–136.

Collentine, J. G. (1997). The effects of irregular stems on the detection of verbs in Spanish. *Spanish Applied Linguistics*, *1*, 3–23.

Collentine, J. (2014). Subjunctive in second language Spanish. In K. L. Geeslin (Ed.) *The Handbook of Second Language Acquisition* (pp. 270–286). John Wiley & Sons, Inc.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, *22*, 1216–1234.

Correa, M. (2011). Subjunctive accuracy and metalinguistic knowledge of L2 learners of Spanish. *Electronic Journal of Foreign Language Teaching*, *8*, 39–56.

Cuetos, F., Glez-Nosti, M., Barbon, A., & Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicologica*, *32*, 133–143.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, *20*, 659–678.

Davies, M. (2006). *A frequency dictionary of Spanish*. Routledge.

Demestre, J., & García Albea, J. E. (2004). The on-line resolution of the sentence complement/relative clause ambiguity: Evidence from Spanish. *Experimental Psychology*, *51*, 59–71.

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish English bilinguals. *Bilingualism: Language and Cognition*, *10*, 101–116.

Felser, C., & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics*, *33*, 571–603.

Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, *34*, 67–98.

Felser, C., Roberts, L., Gross, R., & Marinis, T. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*, 453–489.

Gallego, M., & Pozzi, R. (2018). Mood recognition and production. *Hispania*, *101*, 408–421.

Geeslin, K. L., & Gudmestad, A. (2008). Comparing interview and written elicitation tasks in native and non-native data: Do speakers do what we think they do. In J. Bruhn de Garavito & E. Valenzuela (Eds.), *Selected proceedings of the 10th Hispanic linguistics symposium* (pp. 64–77). Cascadilla Proceedings Project.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498.

Gudmestad, A. (2006). L2 variation and the Spanish subjunctive: Linguistic features predicting mood selection. In C. Klee & T. Face (Eds.), *Selected papers of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 170–184). Cascadilla Proceedings Project.

Gudmestad, A. (2010). Moving beyond a sentence-level analysis in the study of variable mood use in Spanish. *Southwest Journal of Linguistics*, *29*, 25–51.

Gudmestad, A. (2012a). Acquiring a variable structure: An interlanguage analysis of second language mood use in Spanish. *Language Learning*, *62*, 373–402.

Gudmestad, A. (2012b). Toward an understanding of the relationship between mood use and form regularity: Evidence of variation across tasks, lexical items, and participant groups. In K. L. Geeslin & M. Díaz-Campos (Eds.), *Selected Proceedings of the 14th Hispanic Linguistics Symposium* (pp. 214–227). Cascadilla Press.

Gutiérrez, X. (2017). Explicit knowledge of the Spanish subjunctive and accurate use in discrete-point, oral production, and written production measures. *Canadian Journal of Applied Linguistics*, *20*, 1–30.

Hofmeister, P. & Vasishth, S. (2014). Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology*, *5*, 1237.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*, 33–56.

Hopp, H. (2014). Working memory effects on the L2 processing of ambiguous relative clauses. *Language Acquisition*, *21*, 250–278.

Hopp, H. (2016). The timing of lexical and syntactic processes in second language sentence comprehension. *Applied Psycholinguistics*, *37*, 1253–1280.

Hopp, H. (2017a). Individual differences in L2 parsing and lexical representations. *Bilingualism: Language and Cognition*, *20*, 689–690.

Hopp, H. (2017b). Cross-linguistic lexical and syntactic co-activation in L2 sentence processing. *Linguistic Approaches to Bilingualism*, *7*, 96–130.

Hopp. H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, *17*, 5–27.

Hopp, H., & León Arriaga, M. E. (2016). Structural and inherent case in the non-native processing of Spanish: Constraints on inflectional variability. *Second Language Research*, *32*, 75–108.

Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong expectation cancel locality effect: Evidence from Hindi. *PLoS ONE*, *9*: e100987.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–53). Elsevier.

Isabelli, C. A., & Nishida, C. (2005). Development of the Spanish subjunctive in a nine-month study-abroad setting. In D. Eddington (Ed.), *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 78–91). Cascadilla Press.

Iverson, M., Kempchinsky, P., & Rothman, J. (2008). Interface vulnerability and knowledge of the subjunctive/indicative distinction with negated epistemic predicates in L2 Spanish. *Eurosla Yearbook*, *1*, 135–163.

Jegerski, J. (2015). The processing of case in near-native Spanish. *Second Language Research*, *31*, 281–307.

Jegerski, J., & Fernández Cuenca, S. (2019, March). Lexical and morphosyntactic processes in the online comprehension of subject-verb number agreement in L2 Spanish. Paper presented at the Generative Approaches to Second Language Acquisition conference, University of Nevada, Reno, NV.

Jiang, N. & Botana, G. P. (2009, October). Frequency effects in NS and NNS word recognition. Paper presented at the Second Language Research Forum, Michigan State University, East Lansing, MI.

Kaan, E., Ballantyne, J., & Wijnen, F. (2015). Effects of reading speed on second-language sentence processing. *Applied Psycholinguistics*, *36*, 799–830.

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.

Keating, G. D. (2014). Eye-tracking with text. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 69–92). Routledge.

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, *37*, 1–32.

Kimball, A., Shantz, K., Eager, C. & Roy, J. (2018) Confronting quasi-separation in logistic mixed effects for linguistic data: A Bayesian approach. *Journal of Quantitative Linguistics*, *26*, 231–255.

Kirkici, B., & Clahsen, H. (2013). Inflection and derivation in native and non-native language processing: Masked priming experiments on Turkish. *Bilingualism: Language and Cognition*, *16*, 776–791.

Kuznetsova, A., Brockoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). https://cran.r-project.org/web/packages/lmerTest/index.html.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.

Lee, J. (1987). Comprehending the Spanish subjunctive: An information processing perspective. *Modern Language Journal*, *71*, 50–57.

Lemhöfer, K., Schriefers, H., & Indefrey, P. (2014). Idiosyncratic grammars: Syntactic processing in second language comprehensions uses subjective feature representations. *Journal of Cognitive Neuroscience*, *26*, 1428–1444.

Lenth, R., Singmann, H., Love J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated Marginal Means, aka Least-Squares Means. https://cran.r-project.org/web/packages/emmeans/index.html.

Leow, R. (1993). To simplify or not to simplify: A look at intake. *Studies in Second Language Acquisition*, *15*, 333–356.

López Ornat, S., Fernández, A., Gallo, P. & Mariscal, S. (1994). *La adquisición de la lengua española* [The acquisition of the Spanish language]. Siglo XXI.

Lubbers Quesada, M. (1998). L2 acquisition of the Spanish subjunctive mood and prototype schema development. *Spanish Applied Linguistics*, *2*, 1–23.

Massery, L. A. (2009). Syntactic development of the Spanish subjunctive in second language acquisition: Complement selection in nominal clauses [*Unpublished doctoral dissertation*]. University of Florida, Gainesville.

Miller, K. A. (2014). Accessing and maintaining referents in L2 processing of wh-dependencies. *Linguistic Approaches to Bilingualism*, *4*, 167–191.

Montrul, S. (2011). Morphological errors in Spanish second language learners and heritage speakers. *Studies in Second Language Acquisition*, *33*, 163–192.

Montrul, S. & Ionin, T. (2012). Dominant language transfer in Spanish heritage speakers and second language learners in the interpretation of definite articles. *The Modern Language Journal*, *96*, 70–94.

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

Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers: An investigation of the preterite/imperfect contrast in Spanish. *Studies in Second Language Acquisition*, *25*, 351–398.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105–125.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *24*, 501–528.

Pérez-Leroux, A. T. (1998). The acquisition of mood selection in Spanish relative clauses. *Journal of Child Language*, *25*, 585–604.

Pinker, S. (1999). *Words and rules: The ingredients of grammar*. Basic Books.

Pliatsikas, C., & Marinis, T. (2013). Processing of regular and irregular past tense morphology in highly proficient second language learners of English: A self-paced reading study. *Applied Psycholinguistics*, *34*, 943–970.

Poplack, S., Cacoullos, R. T., Dion, N., de Andrade Berlinck, R., Digesto, S., Lacasse, D., & Steuck, J. (2018). Variation and grammaticalization in Romance: A cross-linguistic study of the subjunctive. In W. Ayres-Bennett & J. Carruthers (Eds.), *Manual of Romance Sociolinguistics* (pp. 217–252). De Gruyter.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *85*, 372–422.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Sánchez-Naranjo, J., & Pérez-Leroux, A. T. (2010). In the wrong mood at the right time: Children's acquisition of the Spanish subjunctive in temporal clauses. *Canadian Journal of Linguistics*, *55*, 227–255.

SR Research. (2005). Eyelink 1000 [Apparatus and software]. https://www.sr-research.com/eyelink-1000-plus/

Terrell, T., Baycroft, B., & Perrone, C. (1987). The subjunctive in Spanish interlanguage: Accuracy and comprehensibility. In B. VanPatten, T. R. Dvorak, & J. F. Lee (Eds.), *Foreign Language Learning: A Research Perspective* (pp. 19–32). Newbury House.

VanPatten, B. (1994). Evaluating the role of consciousness in second language acquisition: Terms, linguistic features and research methodology. *AILA Review*, *11*, 27–36.

VanPatten, B. (1996). *Input Processing and Grammar Instruction: Theory and Research*. Ablex.

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, *15*, 225–243.

Wiseheart, M. (2014). Effect size calculator. http://www.cognitiveflexibility.org/effectsize/