

Running head: FINE-TUNING PREDICTIONS IN L2

Note: This manuscript is currently in press at *Studies in Second Language Acquisition*.

The Fine-Tuning of Linguistic Expectations

Over the Course of L2 Learning

Tania Leal*

Indiana University

Roumyana Slabakova

University of Southampton

Thomas A. Farmer

The University of Iowa

*Corresponding Author:

Tania Leal

lealt@indiana.edu

Indiana University

Department of Spanish and Portuguese

Ballantine Hall 844

1020 E. Kirkwood Ave.

Bloomington, IN 47405-7103

Abstract

This study investigates the degree to which native-English speaking learners of Spanish can generate expectations for information likely to occur in upcoming portions of an unfolding linguistic signal. We examine Spanish Clitic Left Dislocation, a long-distance dependency between a topicalized object and an agreeing clitic, whose felicity depends on the discourse. Using a self-paced reading task, we tested the predictions of the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006a,b) and the Reduced Ability to Generate Expectations hypothesis (RAGE; Grüter, Rohde, & Schafer, 2014). Learners successfully demonstrated sensitivity to the violation of expectations set up by the syntactic and discourse context. In addition, the behavior of the L2 learners was dependent on proficiency: the higher their proficiency, the more their behavior mirrored native speaker processing. These results support a view of SLA in which knowledge of L2 discourse-grammatical relationships is acquired slowly over the course of L2 learning.

Keywords: prediction, Clitic Left Dislocation, Spanish, self-paced reading, L2 processing

The Fine-Tuning of Linguistic Expectations Over the Course of L2 Learning

Introduction

Adult native speakers (NSs) of a language possess a wealth of knowledge about how the world works, about how language typically unfolds in the spoken and written modalities, and about the probability with which a linguistic event is likely to occur in different communicative contexts. The utility of this knowledge is foregrounded by the observation that native-speaking readers and listeners use linguistic, visual, and social contexts to generate predictions (and thus pre-activations) for many properties of upcoming linguistic input during on-line comprehension (e.g., Altmann & Kamide, 1999; Arai & Keller, 2013; Bicknell, Elman, Hare, McRae, & Kutas, 2010; DeLong, Urbach, & Kutas, 2005; Dikker, Rabagliati, Farmer, & Pylkkänen, 2010; Farmer, Christiansen, & Monaghan, 2006; Federmeier, 2007; Hale, 2001; Kamide, 2008; Kamide, Altmann, & Haywood, 2003; Kimball, 1975; Levy, 2008; Staub & Clifton, 2006; van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).

Upon encountering an isolated sentence fragment such as, *The conservative political commentator criticized the...*, for example, a reader cannot accurately anticipate the precise word-form that they are likely to encounter next (the context provided by this isolated sentence is not constraining enough), but can generate expectations for, and thus potentially pre-activate representations for, syntactic information (*a noun is likely to appear*), semantic features (*criticize-able things*), and physical features of the upcoming word that are probabilistically related to higher-level expectancies (*noun-like visual features*), among other things. Pre-activating representations for multiple properties of contextually-supported input, even before the input becomes available for sensory processing, likely contributes to the rapidity of real-time language comprehension (e.g., Dikker, Rabagliati, & Pylkkänen, 2009). Knowledge-driven

Once the topicalized phrase (*a Elena*: feminine, singular) has been introduced, an agreeing clitic (*la*: feminine, singular) must subsequently appear in order for the sentence to be grammatical, as indicated by the asterisk in (2). This construction, which must be accompanied by a specific intonation, has different interpretative properties. Note that the dislocated dative NP in (2) is preceded by the preposition “*a*,” a phenomenon that is also referred to as “*a*-marking.” In Spanish, accusative objects are case-marked with this preposition. Crucially for our experimental design, *a* also marks a subset of accusative objects as well, in a phenomenon that is often referred to as “differential object marking” (DOM).¹ Furthermore, CLLD contrasts with another construction in Spanish, Fronted Focus, which involves the dislocation of a phrase not doubled by an agreeing clitic.² The contrasting acceptability of the two is based on different context and intonation: while CLLD is a mark of topicalization, Fronted Focus constitutes a focalization strategy. Thus, evaluating the acceptability of CLLD is not a simple task achieved by following a linear strategy. When Spanish speakers encounter a dislocated, *a*-marked Topic NP, they would expect (predict) that an agreeing clitic should appear downstream. Additionally, they would expect for this clitic to appear pre-verbally, because clitics, being phonologically dependent on verbs, immediately precede finite verbs. In sum, CLLD contains a (non-linear) long-distance syntactic dependency that must be completed, which implies it can be predicted.

¹ To date, there is considerable debate surrounding what defines DOM and DOM objects (e.g., Aissen, 2003; Leonetti, 2004; Torrego, 1998), with notions like specificity and animacy being frequently invoked. Notoriously complex to define, DOM employs dative case markings on some (but not other) direct objects, where these *a*-marked objects are argued to be semantically and pragmatically more salient. According to Torrego (1998), DOM involves both structural and inherent case.

² An example of Fronted Focus is provided below:

- (i) CONTEXT: Do you know at what time they invited Elena to come?
 [A PEDRO] invitaron, no a Elena. No te confundas.
 to Pedro they-invited NEG to Elena. NEG CL confuse.
 “They invited PEDRO, not Elena. Don’t get confused.”

Evidence that this is indeed the case is put forth by Pablos (2006). In an on-line self-paced reading task, she examined patterns of Reading Times (RTs) as native Spanish speakers read sentences that followed a multi-sentence discourse licensing topicalization in two conditions. In one condition, a clitic appeared at the first possible opportunity (before the main verb), conforming to the reader's expectations (Clitic-Present condition). In the other condition, the main verb was not preceded by a clitic, such that the completion of the dependency was delayed until the appearance of a second verb, violating the expectations of Spanish speakers (Clitic-Absent condition).³ Results showed that RTs on the first main verb were found to be significantly higher in the Clitic-Absent condition relative to the Clitic-Present condition. One interpretation of this result is that based on the properties of the discourse and sentential contexts, NSs were able to generate a strong expectation for a clitic to appear before the main verb. When the expected clitic did not appear before the first verb, processing difficulty was higher on the main verb relative to a condition in which the correct clitic appeared before it, and thus when expectations were met. The RT difference across the two conditions can be interpreted, at least in part, as a reflection of the strength of an expectation for a clitic, an expectation that is generated as a function of both knowledge about grammatical relations in one's native language and the context to which speakers are exposed (Hale, 2001; Levy, 2008). Furthermore, we note that the magnitude of the expectation violation effect on the verb could be considered an index of the strength of the knowledge that NSs possess about this grammatical relationship.

The Role of Prediction in Theoretical Accounts of SLA

Although expectancy-driven processing has become a central theme in the on-line native

³ Pablos termed these conditions Clitic-Late and Clitic-Early, respectively.

language processing literature, the degree to which non-native speakers can generate predictions and assess them against incoming linguistic input during on-line L2 processing remains unclear. Presently, there exists considerable evidence within the L2 processing literature that non-native speakers can generate predictions during on-line comprehension in the L2, but that the precision of those expectancies is more variable relative to those of native speakers (e.g., Dallas, DeDe, & Nicol, 2013; Foucart, Martin, Moreno, & Costa, 2014; Grüter, Lew-Williams, & Fernald, 2012; Hopp, 2013; Martin, Thierry, Kuipers, Boutonnet, Foucart, & Costa, 2013). The presence of this variability, we argue, is a point that can serve to inform long-standing debates in the L2 processing literature. Historically, a fraction of the L2 sentence processing literature has focused rather strongly on the differences between the mental representations of grammatical knowledge of native speakers versus those of second language learners. One example of research with such a focus is Clahsen and Felser's (2006a,b) Shallow Structure Hypothesis, which is a model of L2 mental representations that combines linguistic theory and psycholinguistic principles.

The Shallow Structure Hypothesis (SSH) proposes that L2 learners process sentences without mapping detailed syntactic representations because, in order to build meaning, learners can only use lexical-semantic and pragmatic information or strategies that build on templates of argument structure (e.g., agent-verb-patient). Under the SSH, the underdeveloped (i.e., less detailed) grammatical representations of L2 learners are responsible for "shallow" or minimal processing in their second language. Thus, those structures that depend on hierarchical syntactic relations, such as long-distance dependencies (e.g., CLLD) are argued to exist at the most problematic end of the processing difficulty continuum, while adjacent dependencies (e.g., local morpho-syntactic agreement) are argued to posit less difficulty. Crucially, under the SSH, non-

native speakers are not expected to have the ability to generate expectancies about upcoming linguistic material from knowledge bases involving hierarchical syntactic information in the L2.

Recently, in a new proposal, Grüter, Rohde, & Schafer (2014) have put forth another possible explanation. The authors proffer the working hypothesis that the problems arising when non-native speakers process their L2 arise directly from their inability to predict upcoming linguistic material. This working hypothesis is named the Reduced Ability to Generate Expectations (RAGE) hypothesis. It is supported by experimental findings from Korean and Japanese learners of English who were unable to use their knowledge of viewpoint aspect to anticipate the possible referent of ambiguous pronouns (see also Grüter, Rohde, & Schafer, *in press*, for additional data that appear consistent with RAGE). To date, the RAGE hypothesis has only been supported with off-line behavioral data. Grüter and colleagues (2014) used a Truth Value Judgment Task to ascertain whether learners with Korean or Japanese as their native languages knew the meaning of English viewpoint aspect. In a second task, a story continuation task (Rohde, Kehler, & Elman, 2006), the authors provided their learners with sentences such as (3) below and asked them to continue the story as they saw fit. Sentences as in (3a) used the perfective viewpoint aspect, or simple past tense, while sentences as in (3b) used the imperfective aspect, or past progressive tense. Prompts were either free or started with the ambiguous pronoun *he*, which could refer either to the Source argument *Patrick* or to the Goal argument *Ron*.

- (3) a. Patrick_{Source} gave a towel to Ron_{Goal}. (He) _____
 b. Patrick_{Source} was giving a towel to Ron_{Goal}. (He) _____

The difference in the main event viewpoint aspect creates discourse expectations about who would be the subject of the continuation. Consequently, natives continued the imperfective

sentences as in (3b) with a significantly higher percentage of Source continuations. In other words, when the event was presented as incomplete, native speakers thought that *Patrick* was the most likely subject of a follow-up sentence, e.g., *He made sure to give him a clean dry one*. On the other hand, although the L2 learners also exhibited a small preference/bias in favor of the Source argument, their choices of Source continuations were not significantly higher than the Goal continuations. The authors interpreted these findings to signify that “expectation generation at the discourse level is reduced in a non-native language” (p. 190). Thus, the ongoing versus complete situation influenced native expectations about who would be mentioned as the discourse continued. It did not, however, bias the non-native expectations in the same manner.⁴

Both of these accounts (RAGE and the SSH) espouse a worldview in which the L2 linguistic knowledge of a learner—knowledge that must be referenced in order to generate a prediction about the upcoming linguistic signal—is different from that of native speakers. They also propose that these differences should be especially pronounced under certain circumstances. In the case of the SSH, the differences lie in the processing of long-distance dependencies, and in the case of RAGE, they become evident when natives and learners make discourse-level predictions. With respect to on-line processing, these accounts predict that learners and natives will evince different processing patterns on the CLLD manipulation discussed above.

The L2 processing literature indicates that variability in the precision of grammatical predictions is likely attributable, at least in part, to individual differences in a learner’s knowledge of L2 structure—knowledge that is captured in a broad sense by the notion of L2 proficiency (Dussias et al., 2013; Kaan, 2014). Although some evidence from on-line processing experiments has been interpreted as supporting the existence of this qualitative difference (e.g.,

⁴ A major gap in the native and L2 research on this topic (Ferretti et al., 2009) is the lack of theoretical explanation of what exactly creates these discourse biases, and hence, what it is that the learners have to acquire.

Dussias et al., 2013), we note that the effects of proficiency on patterns of L2 processing are mixed. Below, we review some recent evidence that grammatical expectancy effects in on-line adult L2 processing are conditioned upon a learner's level of proficiency. These proficiency effects suggest that the precision of grammatical expectancies in L2 is not stable, but may instead develop gradually and continuously throughout the acquisition process.

Proficiency and Variability in Grammatical Expectations in Learners

To date, a portion of the L2 literature on syntactic predictions has focused on morphosyntactic agreement. A great deal of this research has been conducted by tracking patterns of eye movements to objects in a visual display as participants listen to sentences containing linguistic manipulations. Patterns of looking provide insight into the nature of the representations that listeners activate during comprehension (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Additionally, when participants are able to generate predictions for information likely to be encountered in upcoming portions of the linguistic signal, they often launch anticipatory saccades to objects consistent with their expectations before the predicted information is even heard (e.g., Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003). Offline studies on how well morphosyntactic patterns can be learned by L2ers have produced variable patterns of results (e.g., Franceschina, 2005; Hawkins, 2009; Prévost & White, 2000), although some evidence implicates learner proficiency as one individual difference variable that can account for some of the observed variation.

Research on predictive morphosyntactic agreement has largely focused on determiners. Spanish nouns agree in gender with determiners, which can be either masculine or feminine. Lew-Williams and Fernald (2010) exposed natives and learners to gender-marked determiners

and examined patterns of anticipatory looks to determine whether they could anticipate the upcoming noun from the gender of a preceding determiner. L1 participants were able to anticipate the gender of the noun from the gender of the determiner, although the L2 learners were only able to do so under certain circumstances. Namely, learners were not able to anticipate the gender of known nouns (Exp. 1). In order to determine whether differences in native versus learner frequency of exposure contributed to this difference, they trained all participants on novel nouns (Exp. 2). Learners were able to anticipate the target noun from the gender of the determiner when the same determiner was paired with the same novel noun during both training and testing. They were unable to do so, however, when the training and testing materials used different determiner–noun pairings (Exp. 3). Additionally, the authors observed only a marginal correlation between scores on an L2 proficiency measure and patterns of anticipatory looks in the learners. In a more recent study using the same paradigm, however, Grüter, Lew-Williams, & Fernald (2012) demonstrated that advanced learners were indeed able to generate expectations for the target noun from the gender of the determiner, although they were less consistent in doing so than native speakers. This finding applied to known words and not to novel words, further supporting the idea that, in principle, predicting an upcoming noun based on a gender cue is achievable with sufficient exposure.

In a similar study on determiner noun gender agreement in Spanish, Dussias and colleagues (2013) examined patterns of anticipatory looks in native Spanish speakers and in two different groups of Spanish learners: English-Spanish bilinguals and Italian-Spanish bilinguals.⁵ The English-Spanish group was further divided by proficiency (high, low). Higher-proficiency learners whose L1 was English were able to use gender information predictively during on-line

⁵ This second group was included to determine the effects of transfer, given that Italian has a similar gender marking system to Spanish.

comprehension in a manner quite similar to the native speakers, although less advanced English-Spanish bilingual learners did not appear to do so. The Italian-Spanish group was not divided into proficiency levels although their Spanish proficiency scores were statistically comparable to the intermediate English-Spanish group. Italian-Spanish bilinguals used gender predictively on the feminine but not the masculine items. Taken together, these results suggest that higher proficiency learners may generate expectations that are more native-like in nature even without the benefit of L1 transfer, and that previous experience (in the form of L1–L2 typological similarity) can influence the quality of the expectations generated during processing.

Thus, although many L2 processing studies have demonstrated positive relationships between L2 proficiency and the degree to which learners approximate native-like patterns of on-line comprehension ability (e.g., Dussias et al., 2013; Hopp, 2006; Hoover & Dwivedi, 1998; Jackson, 2008; Lim & Christianson, 2013), the relationship is not always observed. Furthermore, when it is observed, we note that the strength of the reported relationship is quite variable. One possible explanation for the lack of a consensus on the relationship between expectancies in L2 learners and learner proficiency level is that proficiency is often likely to be misrepresented or otherwise underestimated. For example, proficiency is often gauged through self-report measures, and these have been shown to provide inaccurate estimates of the construct (see AlFallay, 2004). Other times, experiments only focus on one very narrow range of proficiency (either beginner, intermediate, or advanced), such that the variability in proficiency values is too low to sufficiently assess the statistical reliability of a proficiency effect. Here, we address these gaps by focusing on the processing of a long-distance dependency in both natives and in learners who span a broad swath of the proficiency spectrum (range 3 - 51, $M = 11.97$, $SD = 8.09$).

Goals of the Present Work

In the work presented here, we investigate whether the precision of the predictions generated during L2 processing is conditioned upon a learner's level of proficiency in the L2. To assess this relationship, we examine how well learners at different stages of L2 learning can generate expectations for a clitic to appear before a main verb when such an expectation is supported by the syntactic and discourse contexts. CLLD is a particularly good structure for testing the development of expectations because it is a long-distance dependency in which the second element can be predicted based on previously occurring syntactic and discourse information. Our work builds on previous offline investigations (Slabakova, Kempchinsky, & Rothman, 2012), which have documented that advanced and intermediate L2 learners are capable of reliably distinguishing CLLD and Fronted Focus—two superficially similar constructions—based on context and intonation. In this study, we examine another aspect of the developing knowledge of this construction through an on-line paradigm, focusing on learners who span a broad swath of the proficiency spectrum. If all L2ers show evidence of qualitatively different (i.e., less detailed) representations (as predicted by the SSH), or of reduced abilities to generate expectations (as predicted by RAGE), we should observe processing patterns that do not conform to NS norms.

In view of this discussion, we put forth two antithetical predictions.

Possibility A. L2 learners will demonstrate behavior consistent with SSH/RAGE.

Evidence supporting this possibility would include data demonstrating differences in processing patterns between the L2 group(s) and the controls when processing CLLD. Namely, L2ers should display differences in their processing patterns in the Clitic-Absent condition vs. the

Clitic-Present condition. In the case of the SSH, this behavior would be attributed to the learners' less detailed syntactic representations. As mentioned earlier, under the SSH, L2 learners are not expected to anticipate syntactic information that depends on long-distance syntactic relationships such as CLLD; they are expected to succeed only if they can use heuristics based on verb frames, and semantic and pragmatic contextual information. According to the SSH, these deficits persist even at very advanced levels such that under this hypothesis, we would not expect to see any developmental progress in learners with proficiency levels ranging from intermediate to advanced.

In the case of RAGE, non-native behavior would be attributed to L2ers' reduced abilities to generate expectations at the discourse level. CLLD depends on discourse context whereby only left-dislocated elements that have been previously mentioned can be resumed by a clitic—when focused, such elements are not clitic doubled. In our experimental materials, crucially, the second part of the long-distance dependency (i.e., a clitic) can only be predicted if speakers are able take into consideration the discourse context. If participants disregard the context, they could attribute the lack of a clitic to another structure (Fronted Focus). In its present formulation, RAGE does not stipulate that reduced prediction abilities are a developmental phenomenon, so L2 proficiency is not predicted to play a role in prediction generation. It should be noted that although RAGE was formulated with prediction at its core, the SSH predates the incorporation of predictive processing in the L2 literature. Nevertheless, Kaan (2014) argues that a predictive component can readily be incorporated into the SSH. Namely, the SSH states that lexical and pragmatic information should be readily available to the L2 speakers, who can then use this information to predict, as long as the dependencies in question are adjacent (e.g.,

upcoming nouns) and not long-distance. Because CLLD involves a long-distance dependency, the SSH predicts that learners should not show native-like predictions of CLLD.

Possibility B: L2ers will demonstrate behavior that is inconsistent with the SSH/RAGE. Under this possibility, the CLLD processing patterns of the L2 groups should become progressively more native-like as scores on an independent proficiency test increase. Additionally, learners with the highest proficiency scores should exhibit native-like patterns of processing with respect to the manipulation. This behavior would challenge the SSH because it would show evidence of native-like processing of a long-distance dependency—processing that is contingent on hierarchical structure. These results would also challenge RAGE because L2 proficiency is not predicted to play a role—L2 speakers are not expected to generate native-like expectations at the discourse level. Results showing evidence of positive development (i.e., non-native-like behavior at the intermediate levels but native-like behavior at the advanced levels) have been claimed to be compatible with accounts that advocate a fundamental similarity of L1/L2 processing (e.g., Hopp, 2013; Kaan, 2014). Under these accounts, using the natives as a baseline, we would expect to see a positive effect of L2 proficiency on the size the expectation violation effects.

Foreshadowing our results, we found evidence that the precision of the expectations generated from syntactic and discourse context in L2 processing increased as a function of proficiency scores. As proficiency increased, L2ers exhibited progressively longer reading times at the main verb when the clitic was absent, relative to when it was present. Moreover, the most highly advanced L2ers and the native speakers were equally sensitive to this effect of expectation violation as evident through highly similar reaction time profiles. This result indicates that both

native speakers and learners possessed potentially very similar knowledge about the grammatical relationship inherent to the CLLD manipulation. Taken together, these results provide partial support for the notion that 1) L2ers are able to generate and assess predictions during on-line comprehension, 2) the precision of the predictions develops progressively over the course of learning, and 3) highly proficient individuals produce patterns of processing behavior that strongly mirror native speakers' upon encountering an expectation violation. In the discussion section we interpret these data in relation to the central claims of the SSH / RAGE hypotheses.

Method

Participants

Second language learners. We tested 120 native English-speaking learners of Spanish (85 women; mean age = 26.4 years, SD = 9.48 years), who were recruited at two universities in the U.S. At the time of testing, participants were either enrolled in undergraduate and graduate Spanish courses or else teaching Spanish courses at the high school or college level. Participants in the L2 groups reported having normal hearing and normal or corrected-to-normal vision. No cognitive disabilities were reported. All the L2 participants were sequential bilinguals who were exposed to Spanish during puberty (mean age of exposure = 14.6 years, SD = 3.34), after they had already learned their first language. None reported fluency in a language other than English or Spanish. Participants received compensation for their participation.

Due to the fact that the CLLD construction is both lexically and syntactically difficult, we aimed to collect data only from learners in the low-intermediate to highly-advanced range of proficiency. Based on proficiency benchmarks articulated in previous literature (e.g., Slabakova, et al., 2012), a learner was considered “advanced” if they produced a score on the independent

measure of proficiency (describe below) between 40–50, and were considered “intermediate” if they scored between 25–39 on the task. Participants with proficiency scores below 25 were considered to have a “beginning” level of proficiency and were excluded from all analyses. Based on this inclusion criterion, 93 of the original 120 learners were included in our sample of learners. Table 1 presents the demographics of the L2 participants, grouped by proficiency level (intermediate vs. advanced), that were ultimately included in the learner sample.

[Insert Table 1 here]

Native speaker control group. Native speakers of Spanish were also tested in order to provide a metric of baseline performance on the processing task against which to assess learner performance. The participants in the Spanish NSs group ($N = 36$, 20 women, mean age = 40.7 years, $SD = 18.7$) were mostly monolingual. They were tested in their native country (Mexico) and reported having normal hearing and normal or corrected-to-normal vision. Spanish NSs reported having had only minimal exposure to English, although all of them had taken at least a year of foreign language (usually English) because at least one year of secondary foreign language education is compulsory in Mexico. Additionally, many of them have been exposed to English through movies, music, and other American cultural products that are ubiquitous in Mexico. Nevertheless, none of the native Spanish participants reported fluency in other languages, including English. In light of the fact that previous research has shown evidence that level of education and other variables such as socioeconomic background are closely related to language outcomes (Mulder & Hulstijn, 2011; Pakulak & Neville, 2010), our NS participants were recruited from a middle-class local arts music conservatory and from the state hospital

(Puebla, Mexico). In order to seek a close match in educational backgrounds, all NS participants had to have completed at least two years of college in order to be included.

Materials

We administered the following four tasks, each detailed below: 1) a language background questionnaire, 2) an independent proficiency test, 3) a test of syntactic knowledge of clitics, and 4) a self-paced reading task designed to gauge the degree of processing difficulty elicited by grammatical expectancy violations in both NSs and learners.

Language background questionnaire. Participants completed this questionnaire in order to collect demographic information, as well as language background (profile of language use) and language learning history.

Proficiency test. In order to determine learners' proficiency level, we administered a test that includes a section of the official language accreditation in Spain (DELE: *Diploma de Español como Lengua Extranjera*), as well as the reading and vocabulary sections of the *Modern Language Association Cooperative Foreign Language Test* (Educational Testing Service). This proficiency test has been used successfully as a discriminator of proficiency in prior L2 Spanish acquisition research (e.g., White, Valenzuela, Kozłowska-MacGregor, & Leung, 2004; Montrul, 2004; Rothman, 2009). Instructions and test items were presented in Spanish. The measure consisted of fifty multiple-choice items focusing on vocabulary and grammar (maximum possible score = 50). The presentation order of the first 30 items was randomized per participant.

The last 20 were part of a coherent paragraph so they were presented in the same order to all participants.

Clitic knowledge test. In order to determine whether the L2 learners had knowledge of the syntactic properties of clitics, participants completed a 10-item multiple-choice task that has previously been used in research on clitic doubling and fronting operations in Spanish with both L2 and Heritage Speaker populations (e.g., Leal Méndez et al., 2014; Slabakova et al., 2012). Ascertaining that our participants had knowledge of the Spanish clitic system was crucial because knowledge of clitics is a natural pre-requisite for learning CLLD. In our task, we probed knowledge of clitic placement rather than agreement.⁶ Romance clitics appear before finite verbs. Additionally, in cases where there is an auxiliary (often a restructuring verb, in the terminology of Rizzi, 1982) and a semantic verb, clitics can either precede the first verb or can attach to the second verb (either a gerund or an infinitive). Each test item included a small dialogue followed by a question, the answer to which required the use of accusative clitics. Every item included two clitics: one accusative and one dative. Clitics have ordering restrictions in Spanish, with dative clitics preceding accusative clitics (Zagona, 2002). Item order and the order of answer choices were randomized per participant. The maximum score for this task was 50 points.

Self-paced reading task. We administered this task to investigate whether participants were able to predict the downstream portion of a non-adjacent dependency based on early-

⁶ Although in our study we only manipulated aspects of clitic placement, rather than agreement, the latter has also been shown to be problematic in L2 acquisition. A recent study by Rossi, Kroll, & Dussias (2014), for instance, showed that advanced L2 learners displayed sensitivity to number violations but not gender violations in the processing of clitic pronouns in the L2.

occurring syntactic cues inherent to the dependency. Stimuli were presented using a non-cumulative word-by-word self-paced moving window display (Just, Carpenter, & Woolley, 1982). Each target sentence was preceded by a one- or two-sentence context that was necessary to make the topicalization natural (i.e., the *topic* was introduced). Each sentence was followed by a yes/no comprehension question. Half of the comprehension questions were based on the context and half on the sentence. For the experimental sentences, the comprehension questions did not focus on clitics.⁷⁻⁸

Two minimally different versions of each test item were presented, corresponding to one of two different conditions: the Clitic Present and Clitic Absent conditions. In the Clitic Absent condition, the sentential-conferred prediction was violated by the omission of the clitic expected to appear before the first verb. In the Clitic Present condition, the predicted clitic was present before the first verb, such that the dependency was completed early. All experimental test sentences were acceptable, but in the Clitic Absent condition, the requirement for a doubling clitic was satisfied later in the structure, thus appearing before the second main verb (ex. 4a). We utilized a modified version of sentence materials originally designed by Pablos (2006) to test long-distance dependencies in a group of Peninsular Spanish NSs. In order to independently verify NSs' preferences, Pablos used an off-line grammaticality-rating task to ensure that NSs preferred clitic over clitic-less sentence continuations after a dislocated phrase was introduced.

⁷ Filler items neither focused on clitics nor included them systematically. Fillers did not involve dislocations of any kind. Clitics are abundant in Spanish and excluding them systematically would be conspicuous.

⁸ We note here that all NSs and all but three learners produced comprehension question accuracy rates above 70%. The results of all analyses reported in the results section are qualitatively the same after excluding the data produced by these three learners.

We modified Pablos' materials by removing one level of embedding and by replacing some lexical items in order to avoid any differences due to dialectal variation.⁹

(3) CONTEXT: Varias estudiantes visitaron el departamento al que habían mandado su inscripción. (*Several students visited the department where they had sent their registration.*)

(4a) Clitic Absent condition (dependency completed late)

Target: **A aquellas estudiantes** la linda secretaria felizmente contó

To those students the lovely secretary happily told.3rd.sg
 que probablemente **las** admitirán en el programa.
 that probably **clitic.ACC** admit.pl.3rd.FUT in the program

(4b) Clitic Present condition (dependency completed early)

Target: **A aquellas estudiantes** la linda secretaria felizmente **les** contó

To those students the lovely secretary happily **clitic.DAT** told.3rd.sg
 que probablemente las admitirán en el programa.
 that probably clitic admit.pl.3rd.FUT in the program

‘The lovely secretary happily told the students that they would probably be admitted to

⁹ We would like to add some remarks about the modification of the items, which involved clitics: Pablos' original study included not two but three clitics in the early condition and two in the late condition. Because our materials could only be considered a partial replication, we focus on the commonalities when reporting Pablos' research which, importantly, had very different goals than ours. To exemplify, we reproduce Pablos' version of (4a) below: (i) Clitic Absent/LATE condition (dependency completed late;)

Target: **A aquellas estudiantes** la secretaria del decano poco después contó

To those students the secretary of-the dean few later told.3rd.sg
 que ya lo espera que **las** admitan en el programa.
 That already CL hopes **clitic.ACC** admit.pl.3rd.SUB in the program

“The dean's secretary, soon after, said that she indeed hopes that these students would be admitted to the program.” The reader is referred to the original source for a more complete—and contextualized—discussion of her dissertation results, which included not one but two self-paced reading tasks.

the program.’

In this manipulation, the sentence-conferred expectation after encountering a left-dislocated phrase (e.g., *A aquellas estudiantes*, in 3a, b) is that an agreeing clitic (agreeing in number and/or gender) must necessarily occur later in the sentence. Given that Spanish clitics must appear before finite verbs (Rizzi, 1982), the clitic would be anticipated to appear immediately before the matrix verb (*contó* in (4)). Following Pablos’s (2006) design, the main verbs in the stimuli were optionally ditransitive (e.g., *decir* ‘to say’). In non-topicalization contexts, these verbs can optionally take a dative argument. This means that this class of verbs can either take a clitic or not ((e.g., ‘say something’ vs. ‘say something to someone’).

Note that the only difference between sentences in each condition is the early vs. late appearance of the doubling clitic. If speakers generated an expectation for an upcoming clitic after encountering a topicalized phrase, they should exhibit processing difficulty (and thus, longer reaction times) upon encountering the main verb when the clitic is absent, relative to when it is present. This raises the point of whether the prediction is local (as a requirement of the verb) or a long-distance one. We will address this issue in the discussion section.

We created two versions of 24 experimental items, one in which the clitic was absent (4a) and one in which it was present (4b). The 48 experimental sentences stemming from 24 items were counterbalanced across two different presentation lists such that each list contained twelve sentences in each condition, but only one version of each item. 48 additional sentences that did not include dislocated items were included as fillers. Participants were randomly assigned to one of two presentation lists, and presentation order was randomized per participant.

Sentence norming task. In order to quantify the degree to which NSs actually expect a clitic to occur before the first verb in the sentence, a separate group of NSs completed a **preliminary normative sentence-completion task**. Monolingual speakers of Mexican Spanish (none of whom participated in any other tasks) completed this task in their native country. The norming task included abbreviated versions of the 24 experimental items used in the self-paced reading task. For each item, participants saw the corresponding contextual information, followed by the fragment of the target sentence up until the first verb (i.e., *A aquellas estudiantes la linda secretaria felizmente ...* from example 4). The rest of the sentence (including the main verb) was replaced by a blank. Participants were asked to complete the sentence quickly, using the first completion that they could think of.

Two versions of this norming task were administered. In one version, participants were given the verb they would have to use in the completion, and in the second version, they were free to use any verb they preferred to use. We constructed two lists in order to ascertain that this preference was not only related to the verbs we used but extended to other verbs. The presentation order of the 24 items, along with 24 intermixed filler items, was randomized per participant. A total of 24 participants completed the first list (verb included), while 25 participants completed the second list (verb open). Sentences were scored by hand and grouped into two categories: (a) sentences were completed using an accusative or dative doubling clitic, or (b) sentences were not completed using a doubling clitic. 87.1% of the sentences in the first list (verb included) were completed using a clitic, and 89.5% of sentences in the second list (verb open) were completed with a clitic. These results show that NSs of Mexican Spanish do have a clear preference for the presence of a doubling clitic after the adverb and thus before the first

verb. This presence was documented both when participants used the same verbs included with the items from our self-paced reading experiment, and also when they were free to choose.

Results

Before data screening or analysis, we removed all RT data from five sentences on which a data recording error occurred (four of which were filler items). We then length-adjusted the raw RT data using a modified version of the length-adjustment procedure recommended by Ferreira and Clifton (1986). This length-adjustment procedure allows us to statistically control for variability in RTs associated with length and overall reading speed on a participant-by-participant basis. Given that our learners were sampled from a broad range of the L2 proficiency spectrum, RTs were likely to be differentially influenced by these variables. We first excluded any data point from either an experimental or filler item with an RT greater than 5000 ms or less than 100 ms. These screening criteria led to the removal of less than 1.5% of the overall number of data points. The remaining raw RTs on each word from every item (both experimental and filler items) were predicted from a regression model that contained a fixed effect for word length, a random intercept for subject, and a random slope for length. The model-predicted values were then subtracted from raw RTs, and these length-adjusted residual RTs served as the dependent variable for all RT analyses reported below (see Fine, Jaeger, Farmer, & Qian, 2013, for a more detailed discussion of the length-adjustment procedure utilized here). We subsequently removed all length adjusted RTs > 2000 ms (only 5 observations).

Linear mixed-effects models were adopted in analyzing length-adjusted RTs. The analyses were implemented with the lme4 package (Bates, Maechler, & Bolker, 2012) in the R environment (R Development Core Team, 2014). Analyses of the length-adjusted RTs were conducted separately for native Spanish speakers and for learners. In all models reported below,

the maximal random-effects structures were utilized (Barr, Levy, Scheepers, & Tily, 2013), always including a random intercept for both participants and items, as well as a random slope for Clitic Condition (Clitic Absent vs. Clitic Present) on both the participant and item terms. Clitic Condition was effect coded (-1 for “Clitic Present” and +1 for “Clitic Absent”). The model for learners included Clitic Condition, Proficiency, and the interaction term as predictors, and the continuous Proficiency variable was centered in order to reduce collinearity between the interaction term and the lower-order effects. Any t -value with an absolute value exceeding 1.96 was considered statistically significant at an alpha level of $p < .05$.

Effect of Clitic Condition on the First (Main) Verb

Natives. Mean RTs on the first verb for each Clitic Condition (Clitic Absent vs. Clitic Present) are presented in Figure 1. A significant effect of Clitic Condition was elicited ($\beta = 95.86$, $SE = 31.14$, $t = 3.08$). When properties of the discourse and sentential context facilitated an expectation for a clitic to occur before the verb, native speakers were significantly faster to read the main verb when it was preceded by the appropriate clitic.

[Insert Figure 1 here]

Learners. Mean RTs on the first verb for each Clitic Condition (Clitic Absent vs. Clitic Present) are presented as a function of a learner’s proficiency score in Figure 2. Table 2 provides a summary of the results of the corresponding statistical model. Significant effects of Clitic Condition and Proficiency were observed. Main verbs were read significantly more quickly when a clitic occurred before them, relative to when a clitic did not, and not unexpectedly,

learners with higher proficiency scores produced faster RTs on the main verb. Crucially, the size of the RT difference between the Clitic Present and Clitic Absent conditions depended on a learner's proficiency score.

[Insert Figure 2 here]

[Insert Table 2 here]

As evident in Figure 2, RTs on the main verb in the Clitic Absent condition increased as proficiency increased, while RTs on the main verb when the expected clitic did appear before it decreased as a function of proficiency. In other words, higher proficiency scores yielded a progressively larger difference in RTs on the first verb when the expected clitic was missing relative to when it was present.

Comparison of Highly Advanced Learners to Natives

To determine whether the effect of Clitic Condition differed between our most advanced learners and native speakers, we selected the adjusted RTs of the learners who scored in the range of 47 to 50 on the proficiency test ($n = 24$, and thus, the top 25.8% of the 93 learners included in our analyses) and compared them to the RTs of the natives with respect to the magnitude of the effect of Clitic Condition. Thus, instead of including continuous proficiency scores in this analysis, we created a new variable Group (Natives, coded as “-1,” versus Highly Advanced Learners, coded as +1).

[Insert Figure 3 here]

A significant effect of Clitic Condition occurred ($\beta = 98.79$, $SE = 12.27$, $t = 6.47$), although no effect of Group (Natives vs. Highly Advanced Learners) occurred ($\beta = 4.63$, $SE = 14.32$, $t = .32$), nor did Group interact with Clitic Condition ($\beta = -3.46$, $SE = 14.00$, $t = -0.25$). As illustrated in Figure 3, RTs in each condition are nearly identical across each group, and the magnitude of the effect of Clitic Condition was not statistically distinguishable between the two groups.

Secondary Analyses on Sentence Regions that Neighbor the First Main Verb

One unavoidable drawback of the CLLD manipulation employed here is that the two conditions differ in the linguistic information a reader must process immediately before encountering the main verb. In the Clitic Present condition, a clitic appears before the main verb, although an adverb appears before the main verb in the Clitic Absent condition. Thus, as participants read the adverb, the possibility exists that they were able to assess, parafoveally, whether or not a clitic was present or absent. This possibility arises for two reasons: 1) by the time that readers encounter the adverb, they will have already processed the information necessary to generate an expectation for a clitic, and 2) the dashes that preserve the spatial layout of the sentence (part of the self-paced reading paradigm) may serve as an informative parafoveal cue that the predicted clitic will or will not appear. Certainly, during natural reading in English, readers can glean visual information from the printed text occurring approximately 4 characters to the left of a current fixation and 14–15 character spaces to the right (e.g., Balota, Polletsak, & Rayner, 1985). This parafoveal information provides a potential source of bottom-up visual information to the reader about the identity of an upcoming word (a short word is likely to be a

clitic, and a longer post-adverbial word is unlikely to be a clitic) before they encounter the target word, even during a self-paced reading task where dashes are displayed but not the actual words.

The possibility that readers were able to parafoveally assess differential length information in the post-adverbial word is potentially problematic for an expectancy-influenced interpretation of the Clitic Condition effects at the main verb, as we report above. In the Clitic Absent condition, we note that our norming data provide compelling evidence that native speakers possess strong expectancies for a clitic upon encountering the adverb. Should readers possess the ability to detect the lack of a clitic through parafoveal processing, they may engage in preprocessing at the adverb that could influence RTs on the clitic (in the Clitic Present condition), or on the target verb in either condition. Evidence of this possibility would manifest itself as an effect of Clitic Condition on the pre-verbal adverb. To assess this possibility, we regressed length-adjusted RTs on the adverb onto Clitic Condition (for the NSs), or onto both Clitic Condition and Proficiency, as well as the interaction term (for the learners), in two additional mixed-effects models. No effect of Clitic Condition occurred for the NSs ($\beta = -53.24$, $SE = 38.81$, $t = -1.37$). For the learners, an expected effect of Proficiency occurred, ($\beta = -10.53$, $SE = 2.25$, $t = -4.69$), such that adverbs were read faster overall as proficiency increased. No effect of Clitic Condition occurred ($\beta = 4.11$, $SE = 17.95$, $t = .229$); however, nor was a Clitic Condition x Proficiency interaction observed ($\beta = 1.53$, $SE = 2.28$, $t = .67$). Thus, participants do not appear to demonstrate sensitivity to the differential visual information that appears to the right of the adverb while reading the adverb.

[Insert Figure 4 here]

Spillover Effect

In self-paced reading experiments, evidence of processing difficulty is often observed not only on a specific word or region where it would be predicted, but also on words occurring after the target region (a “spillover effect”). To determine whether or not the effect of Clitic Condition also exerted an influence on processing of the word appearing after the first main verb, we regressed length-adjusted RTs on the first post-verbal word onto Clitic Condition (for the NSs) or onto Clitic Condition, Proficiency, and the interaction term (for the learners) in two additional mixed-effects models. Length-adjusted RTs on the post-verbal word were not influenced by Clitic Condition for the natives ($\beta = 35.04$, $SE = 20.27$, $t = 1.73$), although we note that the effect is marginally significant. For the learners, however, a significant effect of Clitic Condition did occur on the post-verbal word, such that length-adjusted RTs were higher in the Clitic Absent vs. the Clitic Present condition ($\beta = 23.22$, $SE = 7.56$, $t = 3.07$). No effect of Proficiency was elicited ($\beta = -1.35$, $SE = 0.93$, $t = -1.45$), and crucially, Proficiency did not interact with Clitic Condition ($\beta = 1.05$, $SE = 0.97$, $t = 1.08$). These results suggest that learners may recover more slowly from the processing difficulty experienced upon encountering the main verb without a clitic occurring before it first. To further investigate this potential processing difference, we compared native speaker RTs on the spillover word to the RTs on the spillover word in our most advanced learners ($n = 24$). The analytic strategy was the same as described above in relation to the across-group comparison of RTs on the target verb. We observed a significant effect of Clitic Condition ($\beta = 25.23$, $SE = 9.32$, $t = 2.71$), but no effect of Group ($\beta = 6.89$, $SE = 8.4$, $t = 0.82$), and crucially, no Group x Clitic Condition interaction ($\beta = -8.91$, $SE = 9.32$, $t = -0.96$). Thus, we observed no evidence of a statistically detectable difference in the processing of the spillover word between natives and our highly advanced learners, a pattern borne out in Figure 4.

Clitic Knowledge Task

We demonstrated above that the effect of Clitic Condition in learners is dependent upon Proficiency as gauged by an independent metric. We note here, however, that the Proficiency test we used was not designed to probe knowledge about ways in which clitics can be used in Spanish grammar. Indeed, knowledge of clitic use is crucial to the acquisition of knowledge and accurate comprehension of CLLD. Scores on the proficiency task correlated significantly with scores on the clitic knowledge task, $r = .71$, indicating that as proficiency increased, so did knowledge of clitics. In this sense, the clitic knowledge task can be considered another independent metric of proficiency. Accordingly, we observed that the significant interaction between Proficiency and Clitic Condition reported above also arises when substituting general proficiency scores with scores on the clitic knowledge task ($\beta = 3.97$, $SE = 1.56$, $t = 2.50$).

Quality of Experience Matters More than Quantity

In one final follow-up analysis, we explore the influence of an individual's linguistic experiences on their sensitivity to the lack of the first clitic while using more direct indices of experience. On the language background questionnaire, learners were asked to state the amount of time they had spent learning Spanish, as well as the amount of time that they had studied abroad in a Spanish speaking country. Both estimates were converted into number of months. In separate models, we predicted length-adjusted RTs on the first verb from Clitic Condition and its interaction with either number of months spent studying Spanish, or the number of months spent studying abroad in a Spanish-speaking country. Number of months studying Spanish did not significantly predict RTs on the main verb ($\beta = -0.11$, $SE = 0.10$, $t = -1.12$), nor did it interact

significantly with the already reported effect of Clitic Condition ($\beta = -0.07$, $SE = 0.11$, $t = 0.64$). Number of months spent studying abroad did predict RTs on the main verb ($\beta = -2.31$, $SE = 0.53$, $t = -4.34$), however, and interacted significantly with the effect of Clitic Condition ($\beta = 1.63$, $SE = .62$, $t = 2.61$), which was larger for those who had spent more time in Spanish-speaking countries.

General Discussion

The goal of this study was to determine whether adult second language learners could utilize contextual information to anticipate upcoming grammatical material. The construction employed in our experiment is CLLD, a frequent and salient construction in Spanish, in which a left-dislocated object (a topic) is picked up by an agreeing clitic later in the sentence. English, the learners' native language, also uses fronted objects for the purposes of topicalization, but there are important differences between the Spanish and English topicalizations of objects. First, because English has no clitic pronouns, English topicalized objects are not clitic-doubled.¹⁰ Secondly, as Slabakova (2015) reports, topicalizations are 1,000 times rarer in English than in Spanish, as indicated by corpus data. Thus, although fronted objects are not unattested in English, clitic doubling is something non-native speakers of Spanish have to learn in order to anticipate. Note that in our SPR task, both conditions presented acceptable sentences with fronted objects; however, in the Clitic Absent condition, the expected clitic appeared at the second verb, thus leaving the first verb cliticless, violating expectations.

While prediction plays an important role in L1 processing, the role of prediction in L2 is less clear. Prediction abilities may be reduced because non-native speakers lack robust enough

¹⁰ Instead, there is a null operator in the structure.

knowledge bases (e.g., grammatical knowledge) to generate precise expectancies. As a result of these weakened expectations, non-native speakers are expected to spend more time integrating words into the representation of previously encountered text, or may fail to display the efficiency (and economy of time/effort) that accompanies anticipation. Two well-articulated hypotheses in the SLA literature—the Shallow Structure Hypothesis (C&F) and especially the more recent RAGE hypothesis (G&R)—both maintain that non-native language processing is different from native language processing. The SSH attributes this difference primarily to a lack of native-like knowledge about hierarchical grammatical relations, and thus to an over-reliance on semantic and pragmatic information during processing. In the RAGE hypothesis, the native–learner processing difference is attributed largely to non-native-like processing involving discourse structures.

The SSH and RAGE hypotheses are ripe for further investigation with new language combinations and fresh properties. Filler–gap dependencies have dominated the research in L2 processing (e. g. Dallas, 2008), while long-distance dependencies not involving gaps have been less studied in this context. In this respect, our study helps broaden the scope of long-distance dependency investigation by examining a dependency where the second element is a pronoun rather than a gap. In our experimental setting, as in the one by Grüter and colleagues, we also rely on the discourse expectations of the speakers—the property we examine relates discourse appropriateness and grammatical representations. In this case, by presenting a displaced (topicalized) object, we set up the expectation that a clitic (which must precede a verb) should appear subsequently. Indeed, our norming data demonstrate that native Spanish speakers possess strong expectations for a clitic + verb to appear after the adverb in the materials utilized in our experiment. When the reader’s expectation for a clitic is not met upon encountering a finite verb

instead of a clitic (Clitic Absent condition, 4a), native Spanish speakers spend more time on that verb, signaling that they have noticed the omission and are trying to somehow fit it into the sentence structure. In principle, this is possible in Spanish, as a clitic may appear late, for example, preceding a second verb as in (4b).

As we mentioned earlier, then, the only difference between sentences in each condition is the early vs. late appearance of the doubling clitic. Thus, because participants generate an expectation for an upcoming clitic after encountering the topicalized phrase, processing difficulty (and thus, longer reaction times) upon encountering the main verb should be evidenced when the clitic is absent, relative to when it is present. In this regard, it is relevant to discuss whether this prediction is the result of a long distance dependency or to a local prediction (e.g., due to c-selection or to simple frequency of co-occurrence).¹¹ Pablos (2006) provided evidence that the former is indeed the case. In her dissertation, she included a second experiment in which the conditions included topicalized and non-topicalized versions of the stimuli. The results of the non-topicalized version of the experiment show that the differences in RTs are in fact due to the topicalized element (hence the effects are due to the long-distance dependency and not to a local prediction) because the verbs in both conditions do not show significant RT differences when there is no dislocation (i.e, in non-topicalized sentences). Although these verbs usually have a clitic before them, the participants only displayed higher reaction times in the experiment when the topicalized element appeared, i.e., the cue that triggered the prediction.

Our data are problematic for these hypotheses in their current formulation. First, we find that the learners, similarly to the native speakers, also read main verbs significantly faster when the expected clitic appeared before them. Thus, learner behavior is not qualitatively different

¹¹ We thank an anonymous reviewer for bringing up this point. Admittedly, Pablos demonstrated a long-distance dependence being observed by her native speaker participants. We assume that the same behavior is true of our own native speakers, and compare our learners to those patterns.

from native behavior in this respect. In discussing L2 processing and comparing it to native processing, we are predominantly interested in non-native speakers exhibiting *the same qualitative contrast* as the natives. Specifically, we are looking for a statistical difference between reactions to acceptable as opposed to (temporarily) unacceptable sentence segments. This is precisely the situation we have uncovered in our study. With regard to the SSH, we found evidence that the learners showed sensitivity to the violation of expectations of a structure involving long distance dependencies. In the case of RAGE, we found evidence that our learners displayed sensitivity to the violation of expectations set up by the discourse context (i.e., the discourse licenses a topic, and a left dislocated topic must be resumed by a clitic).

Second, we find that learner behavior was modulated by proficiency. The higher the proficiency score of the learners, the larger the RT difference on the main verb between the Clitic Present and Clitic Absent conditions, as is evident in Figure 2. This result on its own should not come as a surprise to second language researchers. We probed further into the proficiency variable by correlating the RT measures with length of Spanish study and length of study abroad (in months). In principle, the measure of proficiency we used in this experiment, which is very similar to what the majority of Spanish L2A processing studies use, is a type of task that can be perfected during classroom exposure and is not completely immune to metalinguistic rationalization. Thus it is entirely possible that a number of our test participants reached high proficiency without spending significant time exposed to native discourse in a Spanish-speaking country. Length of Spanish study did not correlate with the RT measure of clitic sensitivity in our test. Importantly, length of study *abroad* did correlate with RTs, and did influence the magnitude of the effect of Clitic Condition. This result underscores the fact that the quality of linguistic experience influences the ability to acquire such sensitivity. This observation

is problematic for RAGE and the SSH hypotheses because neither predict that behavior would be modulated with increasing proficiency. Nevertheless, previous research on the SSH (Pliatsikas & Marinis, 2013) has suggested that naturalistic exposure may lead to native-like processing of long-distance dependencies. Our study lends support to this notion.

In fact, other recent psycholinguistic results point to similar conclusions, showing that the quality of the exposure might have primacy over sheer quantity. Davidson (2014), for example, reported a meta-analysis that considered the findings of forty-one ERP studies using logistic regression analyses. The results showed that immersion (closely related to what we report here as study abroad time) was the only factor that significantly impacted the LAN ERP effect, which correlates with syntactic processing mechanisms such as phrase structure violations of the sort we test here. Furthermore, the meta-analysis established that proficiency was also linked with performance, but with the P600 effect, which correlates with later, language-monitoring activities. As mentioned above, studies such as the one conducted by Pliatsikas and Marinis (2013) have also underscored the role of naturalistic exposure in L2 processing, suggesting that this type of exposure can lead learners to overcome shallow processing.

Third, we find that the magnitude of the Clitic Condition effect is *quantitatively* the same across our *most advanced learners* and the native speakers. While not, strictly speaking, necessary, this comparison supports the observation that a great deal of high quality experience with a second language can produce patterns of processing behavior in learners that mirror those produced by native speakers. Additionally, we note that our most advanced learners produced patterns of RTs on the spillover region that are also statistically indistinguishable from native Spanish speakers. These results are consistent with the notion that L2 learners can acquire long-

distance dependencies to native-like levels (contra the SSH) and to master context-dependent predictions (contra the RAGE).

Although our results are problematic for the strongest versions of both SSH and RAGE, as currently formulated, we note that our results by no means invalidate either hypothesis.¹² Self-paced reading data provide an index of processing difficulty. Reaction times do not, however, provide information about the design of the systems that support language learning and processing, the nature of the representations engaged by the system, or the mechanisms that drive incremental interpretation. Thus, while we cannot be certain that the same systems, mechanisms, and representations contributed to the very similar patterns of processing in our most advanced learners relative to NSs, our results are consistent with such a convergence of representations. Additionally, an anonymous reviewer suggests that the fact that the learners display a native-like pattern by the advanced level does not rule out that the mechanisms proposed by these hypotheses could explain the behavior of the intermediate learners. We agree with this point and suggest that these hypotheses could, instead, explain developmental aspects of the L2 learner rather than restrictions on learners' ultimate attainment.

In the context of development, it is important to note that we have not demonstrated that the mechanisms proposed by these hypotheses are not operant during *any* point in development of knowledge about an L2—they merely show that at some point in development, learners are able to shift to a more native-like pattern.¹³ In the case of the SSH, this would mean that learners could, in principle, possess intact representations at some point in development, while for RAGE it would mean that the precision of learner-generated expectancies is not fated to be reduced

¹² We note that although our data does not provide support for either SSH or RAGE, we do not claim to be able to adjudicate between the two with our current methodology.

¹³ We thank an anonymous reviewer for raising this issue.

permanently. Indeed, modifications of either hypothesis could be instituted to accommodate our data (e.g., Kaan, 2014).¹⁴ For example, SSH could be modified to postulate a progressive change in the types of information that adult second language learners rely on as a function of increasing proficiency. It may be the case that lexical, semantic, and pragmatic information sources are important at earlier points in time during the development of L2 knowledge bases, but that L2 grammatical knowledge bases may develop and guide processing at later points in L2 learning. With regard to RAGE, it may be that prediction abilities are dependent on the availability of certain (linguistic) pre-requisites. It seems logical to assume that prediction is heavily dependent on lexical knowledge, for example, and indeed there is some evidence that this is the case (Hopp, 2013). It may also be the case that learners have to accrue a minimal threshold of syntactic knowledge before having the ability to predict, although this would have to be shown experimentally.¹⁵ More research on these hypotheses could uncover potentially crucial knowledge about how these mechanisms shape and constrain development.

Overall, our results demonstrate that incrementally increasing amounts of exposure to the L2 gives rise to more detailed knowledge of L2 syntactic relationships in learners, thus contributing to the generation of increasingly more native-like syntactic expectations across the trajectory of development. The unavoidable question, then, is *how*, exactly, predictions develop over the course of acquisition—that is, how does this knowledge come to exist in the first place? This is an open question, given that the development of predictions over the course of learning (either in L1 or L2) is not an issue that has been resolved at the present time in the processing literature. With regard to L1 processing, although it is often noted that children appear to learn

¹⁴ We note here that neither SSH or RAGE were intended to be models of L2 development.

¹⁵ We note here that although they do not re-formulate the RAGE hypothesis, Grüter et al. (*in press*) have recently noted that learners may use different strategies—one of which involves linguistic pre-processing—at different stages of L2 learning. This observation is consistent with the possibility that RAGE can be further refined to include a stronger developmental component.

language in a quick and relatively effortless manner (e.g., Pinker, 1995) with surprisingly few “logically possible errors” (Snyder, 2007), child acquirers display a host of well-documented individual differences in the developmental trajectories underlying many cognitive abilities (e.g., Bates, Bretherton, & Snyder, 1988; Zangl, Klarman, Thal, Fernald, & Bates, 2005). Individual differences in child L1 vocabulary knowledge have been linked to variability in many factors that affect language processing, including phonetic discrimination (e.g., Werker, Fennel, Corcoran, & Stager, 2002), as well as the processing of a degraded linguistic signal using verb-based semantic information (Zangl et al., 2005).

In L1 acquisition, predictions are likely to emerge gradually, alongside the learning/acquisition of linguistic properties of a learner’s environment, knowledge about objects and events that occur in the world, and the development of other social and cognitive abilities. Not surprisingly, given the variety of potential factors involved, there are many open questions regarding how these cognitive skills influence child L1 linguistic development (Weisleder & Fernald, 2013). In this respect, the investigation of the development of predictions in adult L2 may offer unique advantages because adult learners are mostly “done” with social-emotional and cognitive development by the time they encounter the L2. Thus, the investigation of the incremental formation of linguistic expectations in adult L2ers is likely to serve as one important gauge of L2 development, but it may also have consequences for our understanding of the contributions of learning to development of expectations in L1.¹⁶

With respect to the development of predictions in L2 acquisition, Kaan, Dallas, and Wijnen (2010) have suggested that the answer involves the “noticing” of L2 predictive cues.

¹⁶ Although our study does not focus on this issue, we note that a crucial aspect of the future investigation of the incremental formation of linguistic expectations should take into consideration individual differences, which could likely affect any shifts in the developmental continuum. We thank an external reviewer for her/his contribution on this important point.

Kaan and colleagues propose that these cues can be noticed differentially based on properties of the linguistic characteristics of the L1 and the L2, such as L2 frequency of occurrence and the extent of featural overlap between the two languages. Arguably, once learners “notice” a cue, they can determine the probability of a given continuation (i.e., predict). The development of predictions is supported by learning from the difference between the predicted and the actual outcome (i.e., the error), which can be used to adjust the weighing of likelihoods associated with predictive relationships (Kaan, 2014). Once this process of adjustment undergoes numerous cycles, learners are expected to generate precise predictions in real time, just like L1 speakers (Kaan et al., 2010; Kaan, 2014). Under this worldview, noticing errors is crucial for the development of predictions in an L2. Therefore, learners should become increasingly sensitive to errors/violations over the course of learning, facilitating stronger knowledge of grammatical relations in an L2 over time. An open question on this topic is whether all L2 learning in adult learners is then error-driven. This question should be addressed experimentally.

As we mentioned earlier, the findings from our cross-sectional data show evidence of positive development across proficiency levels. Historically, the actual progression of interlanguage development has been viewed differently in different SLA paradigms—especially between generative and usage-based SLA. Here, however, we concentrate on a point of agreement between these two paradigms: the primacy of the input in the acquisition process. CLLD, a relatively transparent and ubiquitous construction, does not constitute a poverty of the stimulus situation. As such, our data do not allow us to tease apart whether the effects of increased exposure could be aided by universal principles, by some other domain-general mechanisms such as those that underlie statistical learning ability, or by an interaction between both. Within generative approaches, interlanguage development is limited by those options that

are available in all human languages (Universal Grammar). L2ers are hypothesized to know the properties and syntactic consequences of functional categories although they have to learn how these are realized in the L2. This means there is a great deal that has to be learned, possibly by means of error adaptation or associative learning. Our data demonstrate that with enough high quality experience and linguistic input, L2 learners can produce patterns of processing behavior that appear native-like. We note here, however, this result might not hold for the acquisition of other linguistic properties in the L2.

Conclusion

The current study investigated the on-line generation of expectations by examining patterns of processing of CLLD, a long-distance dependency at the syntax-discourse interface, in a group of Spanish NSs and a group of L2 learners spanning a particularly broad swath of the proficiency spectrum. We examined the predictions of the SSH (Clahsen & Felser, 2006) and the RAGE hypotheses (Grüter, Rohde, & Schafer, 2014), which predict asymmetries between NS and L2 participants with regard to the generation of expectations in this long-distance dependency. Results showed that our L2ers successfully demonstrated an emergent sensitivity to the violation of expectations set up by the syntactic and discourse context. Moreover, their behavior was dependent on proficiency: the higher the proficiency, the larger the RT difference at the region of interest when contextually-conferred expectancies were satisfied as opposed to when they were violated. At the highest level of proficiency, a subset of our most advanced learners produced an RT profile that was not statistically distinguishable from that of NSs. Additionally, we observed a correlation between the magnitude of the RT difference and study abroad. We argued that this

result underscored the notion that the quality of linguistic experience can influence the ability to generate more precise expectations during on-line processing.

References

- Aissen, J. (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory*, 21, 435-483.
- AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self-and peer-assessment. *System*, 32(3), 407-425.
- Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247-264.
- Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, 28(4), 525-560.
- Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive psychology*, 17(3), 364-390.
- Bates, E., Bretherton, I., Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.
- Bates, D., Mächler, M. & Bolker, B. (2012). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 55, 1-9.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.

- Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, *63*(4), 489-505.
- Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.
- Clahsen, H., & Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*, 564–570.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181-204.
- Dallas, A. C. (2008). *Influences of verbal properties on second-language filler-gap resolution: A cross-methodological study*. Unpublished doctoral dissertation, University of Florida.
- Dallas, A. C., DeDe, G., & Nicol, J. (2013). An Event-Related-Potential (ERP) investigation of filler-gap processing in native and second language learners. *Language Learning*, *63*, pp. 766-799.
- Davidson, D. (2014). Electrophysiological changes during word or grammar learning and the role of feedback. Plenary talk presented at the *Second Language in the Brain* symposium, University of Greenwich, Oct 4th, 2014.
- Dell, G., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394.
- DeLong, K., Urbach, T., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117-1121.

- Dikker, S., Rabagliati, H., Farmer, T. A., & Pytkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science, 21*(5), 629-634.
- Dussias, P., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand. *Studies in Second Language Acquisition, 35*(02), 353-387.
- Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*(32), 12203-12208.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*, 491-505.
- Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of memory and language, 25*(3), 348-368.
- Ferretti, T. R., Rohde, H., Kehler, A., & Crutchley, M. (2009). Verb aspect, event structure, and coreferential processing. *Journal of Memory and Language, 61*, 191-205.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS One, 8*(10), 1–18.
- Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), pp. 1461-1469.
- Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender*. Amsterdam: John Benjamins Publishing.
- Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research, 28*(2), 191–215.

- Grüter, Theres, Hannah Rohde & Amy J. Schafer (*in press*). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*.
- Grüter, T. Rohde, H., & Schafer, A. J. (2014). The Role of discourse-level Expectations in non-native speakers' referential choices. *BUCLD 38 Proceedings* (pp. 179-191). Somerville, MA: Cascadilla Proceedings Project.
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies* (pp. 1-8). Assoc. for Computational Linguistics.
- Hawkins, R. (2009). Second language acquisition of morphosyntax. In Ritchie, W., Baktia, T. (Eds.), *The New Handbook of Second Language Acquisition* (pp. 211-233). Leeds: Emerald.
- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22(3), 369-397.
- Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29(1), 33-56.
- Hoover, M. L., & Dwivedi, V. D. (1998). Syntactic processing by skilled bilinguals. *Language Learning*, 48(1), 1-29.
- Jackson, C. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, 58(4), 875-909.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of experimental psychology: General*, 111(2), 228.

- Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different?. *Linguistic Approaches to Bilingualism*, 4(2), 257-282.
- Kaan, E., Dallas, A. C., & Wijnen, F. (2010). Syntactic predictions in second-language sentence processing. In J.-W. Zwart, & M. de Vries (Eds.), *Structure preserved. Festschrift in the honor of Jan Koster* (pp. 207–213). Amsterdam: John Benjamins.
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, 2(4), 647-670.
- Kamide, Y., Altmann, G., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133-156.
- Kempchinsky, P. (2013). CLLD as a Window on the Left Periphery. In C. Howe, S.E. Blackwell, & M. Libbers Quesada, *Selected Proceedings of the 15th Hispanic Linguistics Symposium* (pp. 310-327). Somerville, MA: Cascadilla Proceedings Project.
- Kimball, J. (1975). Predictive analysis and over-the-top parsing. In J. Kimball (Ed.), *Syntax and Semantics*, 5. New York: Academic Press.
- Lim, J. H., & Christianson, K. (2013). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, 16(03), 518-537.
- Leal Méndez, T., Slabakova, R., & Rothman, J. (2014, in press). Discourse-sensitive Clitic-doubled Dislocations in Heritage Spanish. *Lingua*.
- Leonetti, M. (2004). Specificity and differential object marking in Spanish. *Catalan Journal of Linguistics*, 3(2004), 75-114.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.

- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63, 447–464.
- Martin, C., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, 69(4), 574–588.
- Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism Language and Cognition*, 7(2), 125-142.
- Mulder, K., & Hulstijn, J. H. (2011). Linguistic skills of adult native speakers, as a function of age and level of education. *Applied linguistics*, 32(5), 475-494.
- Pablos, L. (2006). *Pre-verbal structure building in Romance languages and Basque*. Unpublished dissertation. The University of Maryland.
- Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, 22(12), 2728-2744.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(04), 329-347.
- Pinker, S. (1995). Language acquisition. In L. Gleitman, & M. Liberman (Eds.), *Language: An invitation to cognitive science*, (pp. 135-182). Cambridge, MA: MIT Press.
- Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, 16(1), 167-182.

- Prévost, P., & White, L. (2000). Truncation and missing inflection in second language acquisition. In: Friedmann, M.A., Rizzi, L. (Eds.), *The Acquisition of Syntax* (pp. 202–235). London: Longman.
- R Development Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rizzi, L. (1982). *Issues in Italian syntax*. Berlin: Walter de Gruyter.
- Rohde, H., Kehler, A., & Elman, J. L. (2006). Event structure and discourse coherence biases in pronoun interpretation. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 617-622). Mahwah, NJ: Lawrence Erlbaum.
- Rothman, J. (2009). Pragmatic deficits with syntactic consequences?: L2 pronominal subjects and the syntax–pragmatics interface. *Journal of Pragmatics*, 41(5), 951-973.
- Slabakova, R. (2015). The effect of construction frequency and native transfer on second language knowledge of the syntax–discourse interface. *Applied Psycholinguistics*, 36, 3, 671–699.
- Slabakova, R., Kempchinsky, P., & Rothman, J. (2012). Clitic-doubled left dislocation and focus fronting in L2 Spanish: A case of successful acquisition at the syntax–discourse interface. *Second Language Research*, 28(3), 319-343.
- Snyder, W. (2007). *Child language: The parametric approach*. Oxford University Press.
- Rossi, E., Kroll, J. F., & Dussias, P. E. (2014). Clitic pronouns reveal the time course of processing gender and number in a second language. *Neuropsychologia*, 62, 11-25.

- Staub, A., & Clifton Jr., C. (2006). Syntactic prediction in language comprehension: Evidence from either... or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 425.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634.
- Torrego, E. (1998). *The Dependencies of Objects*. Cambridge, MA: MIT Press.
- Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443.
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143-2152.
- Werker, J. F., Fennell, C.T., Corcoran, K.M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, 3(1), 1-30.
- White, L., Valenzuela, E., Kozłowska-Macgregor, M., & Leung, Y. K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25(01), 105-133.
- Zagona, K. (2002). *The syntax of Spanish*. Cambridge: Cambridge University Press.
- Zangl, R., Klarman, L., Thal, D. J., Fernald, A., & Bates, E. (2005). Dynamics of word comprehension in infancy. *Journal of Cognition and Development*, 6(2), 179-208.

Table 1

Means (SDs) for the demographic information of the L2 Learners, by group, who satisfied the criterion for inclusion.

	Advanced	Intermediate
N	56	37
Women	37	24
Age	32.5 (10.5)	21.3 (3.6)
Years of exposure (mean)	14.9 (8.8)	7.5 (2.8)
Proficiency scores (mean)	45 (31)	31 (4.6)

Table 2

Result of the LME on length-adjusted RTs elicited by the main verb for all learners. $|t's| > 1.96$ are considered statistically significant at an alpha level equal to .05.

	β	SE	t
Intercept	-68.19	10.65	-6.40
Clitic Condition	68.35	11.42	5.99
Proficiency	-3.97	1.26	-3.16
Clitic Condition *			
Proficiency	3.99	1.33	3.00

Figures

Figure 1. *Length-adjusted Reaction Times for Native speakers in the Clitic Absent (expectation violated) and Clitic Present (expectation met) conditions.*

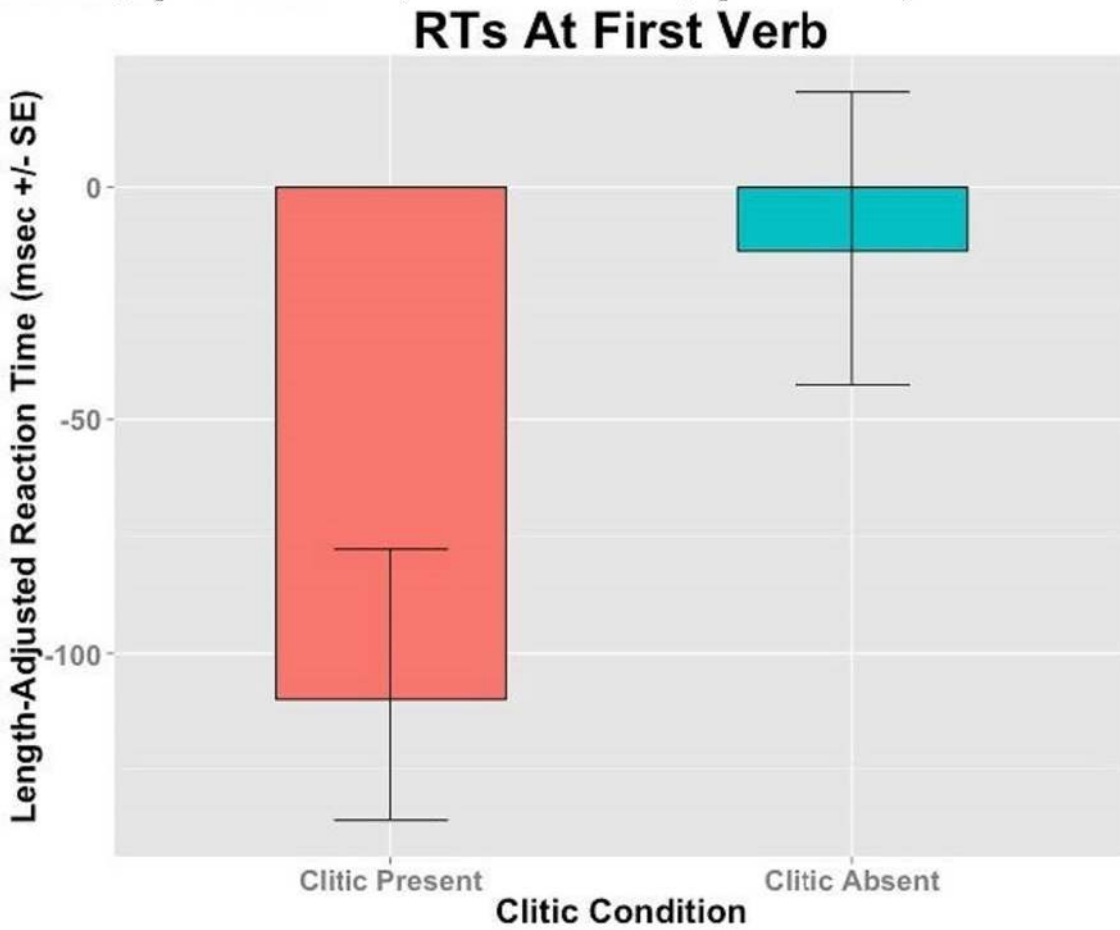


Figure 2. L2 learners: Difference in RTs at each level of Clitic Condition as a function of Proficiency.

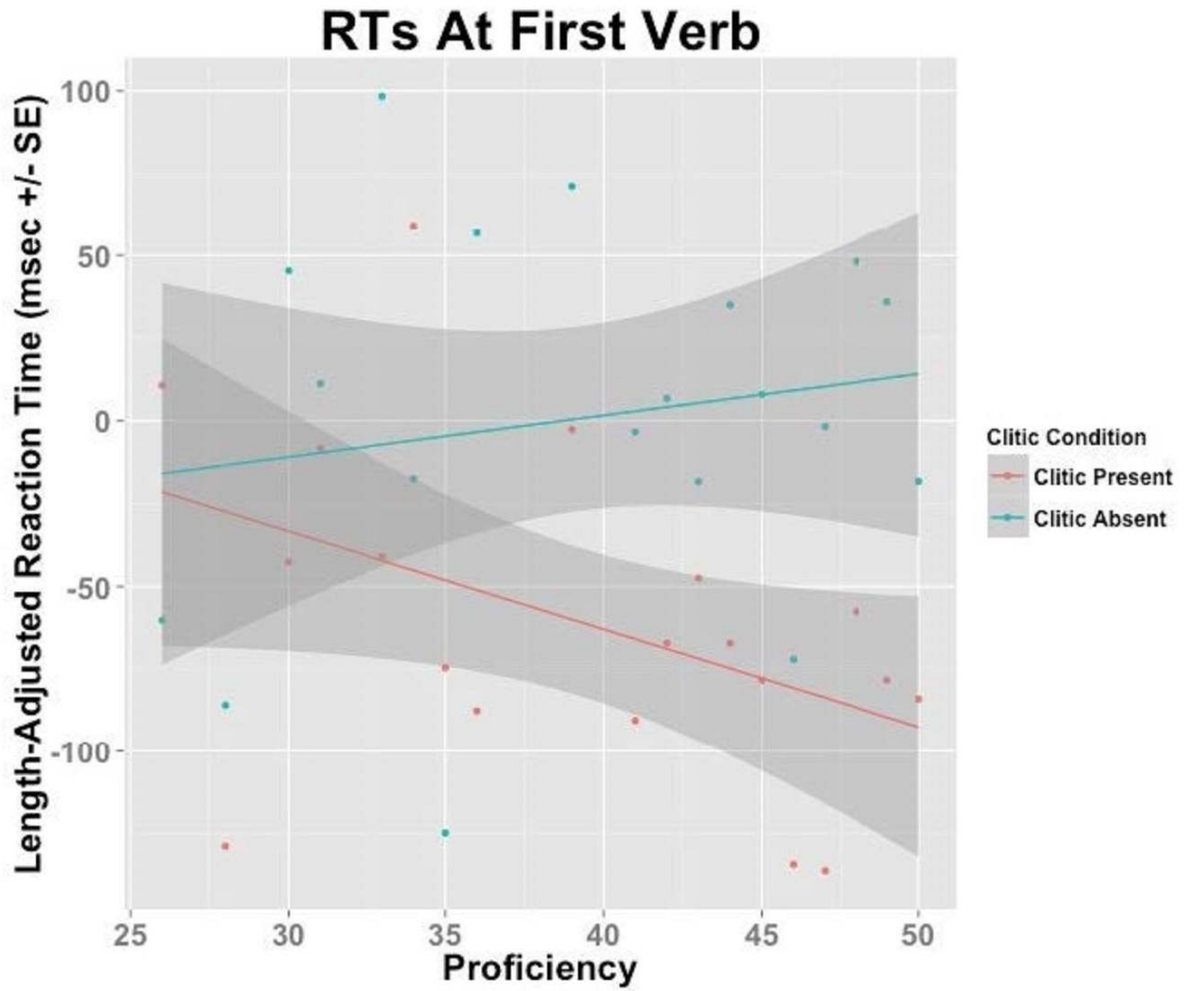


Figure 3. By-condition *length-adjusted RTs* for *NSs* and *highly advanced L2ers* on the main verb.

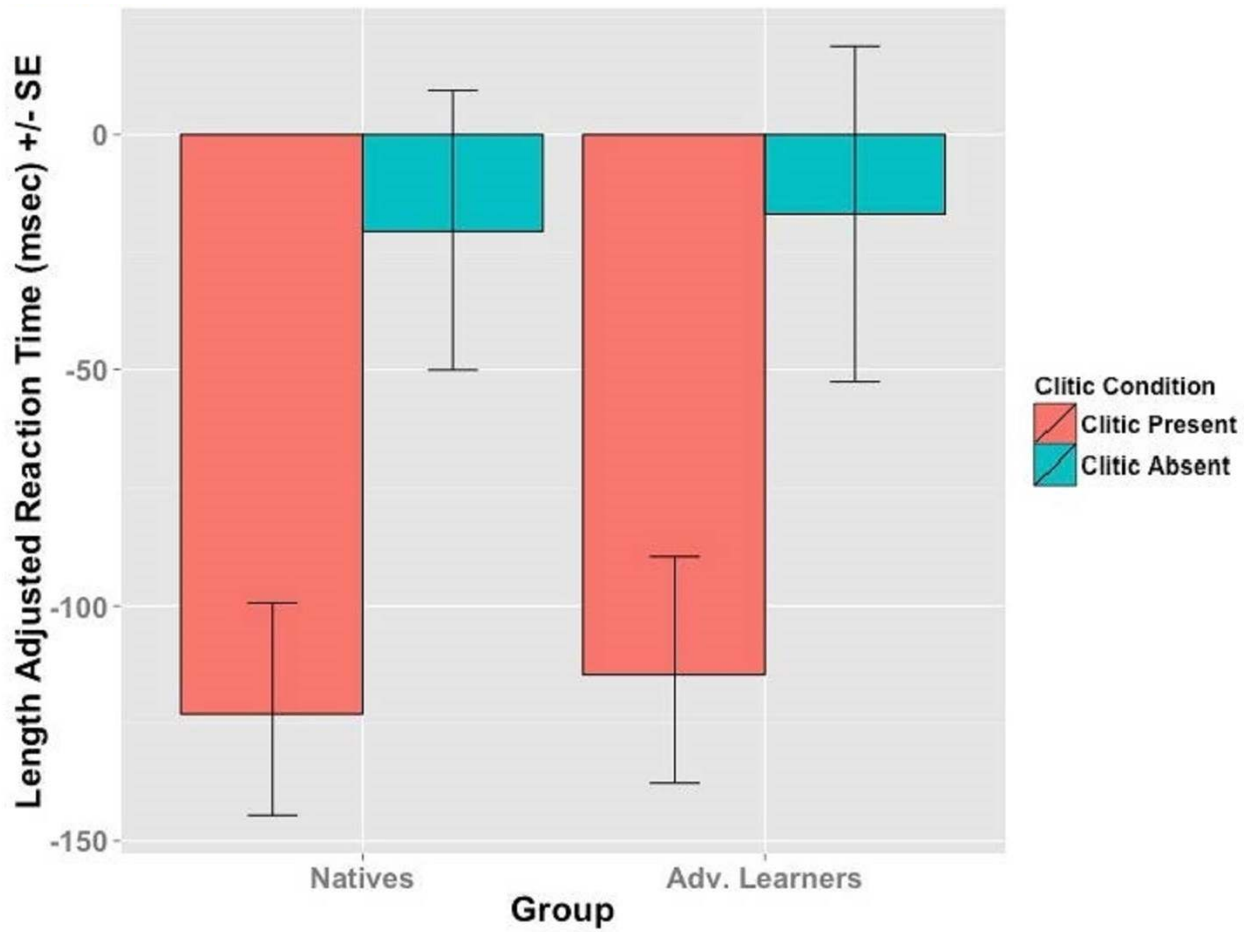


Figure 4. By-condition *length-adjusted RTs* for *NSs* and *highly advanced L2s* on the word appearing immediately after the target verb (*spillover region*).

