Second Language Processing Efficiency: Experience and Cognitive Effects on L2 Morphosyntactic Integration and Anticipation

Crystal Marull

1. Introduction to the Study

The purpose of this study was to test the RAGE hypothesis and to tease apart L2 predictive processes from integrative processes. Native speakers develop integrative and predictive mechanisms early on, but second language (L2) learners often fail to develop native-like processing. Current L2 studies point to morphosyntactic anticipation as the source of processing breakdown even when learners have extensive language exposure and high levels of proficiency (e.g., Dussias et al., 2013; Grüter, Lew-Williams, & Fernald., 2012, Hopp, 2013, 2014, 2015; Lew-Williams & Fernald, 2007, 2010). However, there has not been a systematic analysis of the processor’s subcomponents in the L2 literature. Instead researchers have preferred to approach processing in holistic terms. Teasing apart these two processes contributes significantly to current debates on the locus of difficulty in L2 sentence processing and has the potential to advance knowledge and understanding of what non-native processing exactly entails.

According to the RAGE hypothesis, learners are expected to be capable of native-like integration of morphosyntax, but are expected to show reduced ability to generate expectations based on morphosyntactic cues. Additionally, cognitive resources are expected to play a central role such that learners will be more likely to show native-like processing when the constructions share greater L1/L2 similarities and/or when learners have increased proficiency.

The series of experiments advanced in this study test (a) L2 learner ability to integrate number agreement in real time, and (b) explore L2 learner ability to use morphosyntactic number cues to generate expectations about upcoming input in real time.

The study aims to answer the following specific research questions:

1. Are intermediate and advanced L2 learners of Spanish sensitive to number violations between demonstrative-noun and definite article-noun constructions during reading of L2 Spanish?

* Crystal Marull, Rutgers University, crystal.marull@rutgers.edu. Thank you to everyone who made this study possible and for the support from the Mellon Summer Study Grant and multiple Pre-dissertation Research Awards from Rutgers University.

2. Are intermediate and advanced L2 learners of Spanish able to use number information from the demonstrative and the definite article to anticipate an upcoming target object to improve accuracy and response time on a picture disambiguation task?

2. Background

The key to achieving native-like processing in the L2 may lie in the learner’s ability to develop specific and distinct strategies to overcome the cognitive demands at different levels of processing. This study aims to elucidate the differential effects the learner’s proficiency and language experience (both in the L1 and L2) on two specific subcomponents of the processor. Motivated by the breadth of related L1 research and the relative dearth of L2 research, the two processing subcomponents selected for investigation in this study are integration and anticipation. Integration is a bottom-up reactive mechanism that assimilates incoming information with previously encountered information in a serial fashion (e.g., Frazier, 1999; Kazanina et al., 2007; Phillips & Ehrenhofer, 2015; Traxler & Pickering, 1996). In contrast, anticipation is a top-down mechanism that proactively generates expectations about upcoming material based on previously encountered cues that facilitate integration of bottom-up input further down in the processing pipeline (Borovisky et al., 2012; Elman, 1990; Federmeier, 2007; Misyak, Christiansen, & Tomlin, 2010).

In previous L2 studies learners have been successful in achieving native-like abilities, for example, on comprehension tasks involving (ungrammatical) gender-mismatch between nouns and post-nominal adjectives, case marking, and verb semantics (e.g. Dowens, Vergara, Barber, & Carreiras, 2009; Keating, 2009; Hopp 2010). Nonetheless, preliminary evidence shows that even highly proficient L2 learners differ from native speakers on their ability to exploit morphosyntax to anticipate upcoming constituents to predict or facilitate integration of the upcoming constituent (Grüter, Lev-Williams, & Fernald, 2012; Lew-Williams & Fernald, 2010; Martin et al., 2013; Kaan, 2007; Hopp, 2015). Within the L1 literature, anticipation is widely accepted as being fundamental to driving incremental processing in sentence comprehension (Altmann & Mirkovic, 2009; Federmeier, 2007; Gibson, 1998; Kamide, 2008; Levy, 2008; Pickering & Garrod, 2013).

Therefore, if this mechanism is inefficient in the L2 it will have a negative impact on sentence comprehension.

Taken together, the previous findings begin to paint a picture whereby L2 processing difficulties may be localizable, not in the bottom-up integration mechanisms, but rather in the inefficiency of the anticipatory mechanism to generate linguistic expectations based specifically on morphosyntactic features. In this current study, the terms “prediction,” “anticipation,” and “expectation” are used interchangeably. Furthermore, I adopt Kuperberg and Jaeger’s (2016) definition that these terms minimally refer to the notion that contextual information (from multiple sources) is used to change the state of the language processing system before
new bottom-up input becomes available, thereby facilitating processing of this new input.

Most current models of language processing posit that the comprehender employs a predictive strategy to anticipate some syntactic or semantic information prior to the unfolding of the bottom-up information of such constituents (Kuperberg & Jaeger, 2016). What remains controversial is the scope and nature of the mechanism. Thus, this study tests the assumptions and predictions of the RAGE hypothesis to identify a context in which native-like integrative and anticipatory processing is possible. If evidence can be found that the L2 learner is capable of native-like morphosyntactic anticipation under any circumstances, then it will be most parsimonious to assume that the L2 processor is not inefficient or underdeveloped, but rather, influenced by interactions between cognitive abilities and language experience.

3. Research Design

This study tests Spanish natives (recruited in Argentina) and English learners of L2 Spanish (recruited at a large public institution) on their ability to exploit linguistic cues to anticipate upcoming information. All participants were between 18 to 45-years-old and, minimally, had completed high school. A total of 100 participants were recruited to participate in this study: Spanish native speakers (N=32), Advanced English-Spanish learners (N=19), and Intermediate English-Spanish learners (N=49).

The learner participants completed a series of linguistic tasks completed in two sessions on separate days, a pre-session lasting approximately 35 minutes and the experimental session lasting approximately 60-90 minutes. The natives completed all tasks, including those of the pre-session, in a single session of 60-90 minutes. All materials were administered via digital presentation on a computer screen which required participants to read, listen, speak, and click to answer. The experimental design is shown below in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Experimental Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Session (Learners only):</td>
</tr>
<tr>
<td>Language Questionnaire (5 minutes)</td>
</tr>
<tr>
<td>Experimental Session (All):</td>
</tr>
<tr>
<td>Self-paced reading task (20 min)</td>
</tr>
</tbody>
</table>

Note. The presentational order of the self-paced reading and picture-selection task was counterbalanced across participants such that half of the participants completed the self-paced reading task prior to the picture-selection task and the other half completed them in reverse order.
3.1. Materials and Procedure

3.1.1. Pre-screening Questionnaire and Proficiency Test

The participants completed a language background questionnaire to determine linguistic background, dominant languages, and previous and current contact with Spanish and/or other languages. No participant was excluded based on the information collected in this questionnaire.

Both Native and Learner participants also completed a Spanish proficiency test - a modified version of the Diploma de Español como Lengua Extranjera (DELE) (Diploma of Spanish as a Foreign Language). The scores on this test were used to divide the latter group into advanced and intermediate proficiency groups (Natives: \( M = 28.34, SD = 2.86 \); Advanced: \( M = 25.63, SD = 3.4 \); Intermediate: \( 15.6, SD = 2.63 \)).

3.1.2. Stimuli for the Self-Paced Reading and Picture-Selection Task

The critical sentences, adapted and modified from Tokowicz and Warren (2010) for the reading task, manipulated number agreement between the determiner and noun. The number of the determiner (singular plural) was counterbalanced across all trials. Crosslinguistic similarity (demonstratives or definite article) and grammatical number agreement in the determiner phrase (agreement or violation) were manipulated resulting in the following four conditions:

(1) Demonstrative agreement
Por la tarde el periodista entrevistó a esos\textsubscript{pl} payasos\textsubscript{pl} en el circo.
‘In the afternoon the journalist interviewed those\textsubscript{pl} clown\textsubscript{pl} at the circus.’

(2) Demonstrative violation
Por la tarde el periodista entrevistó a esos\textsubscript{pl} payasos\textsubscript{sg} en el circo.
‘In the afternoon the journalist interviewed those\textsubscript{pl} clown\textsubscript{sg} at the circus.’

(3) Definite article agreement
Por la tarde la periodista entrevistó a los\textsubscript{pl} payasos\textsubscript{pl} en el circo.
‘In the afternoon the journalist interviewed the\textsubscript{pl} clowns\textsubscript{pl} at the circus.’

(4) Definite article violation
Por la tarde la periodista entrevistó a los\textsubscript{pl} payasos\textsubscript{sg} en el circo.
‘In the afternoon the journalist interviewed the\textsubscript{pl} clown\textsubscript{sg} at the circus.’

The linguistic stimuli for the picture task was the same as the self-paced reading task, except that only grammatical sentences were used and the master sentence was manipulated for crosslinguistic similarity and cue informativity to create the following four conditions: demonstrative informative (number mismatch in the visual stimuli), demonstrative uninformative (number match in the visual stimuli), definite informative (number mismatch in the visual stimuli), or definite
uninformative (number match in the visual stimuli). For example, in the informa-
tive condition, participants heard a sentence like, “Por la tarde el periodista en-
trevistó a esos payasos en el circo,” (English: In the afternoon the journalist in-
terviewed those clowns in the circus) and both images contained a journalist but
in the target image he is interviewing two clowns and in the distractor image he is
interviewing just one magician (mago). Thus the plural number marking on the
determiner is informative to initiate disambiguation between the two images prior
to the unfolding of the bottom-up lexical content of the noun.

In contrast, in the demonstrative uninformative condition, participants that
heard the same sentence “Por la tarde el periodista entrevistó a esos payasos en
el circo,” (English: In the afternoon the journalist interviewed those clowns in the
circus) saw a target image in which the journalist is interviewing two clowns and
in the distractor image in which he is interviewing two magicians (magos). Thus,
plural number-marking is not informative to initiate disambiguation between the
two images and participant must wait for lexical content of the target noun to
proceed with disambiguation. Gender was held constant (masculine), and the im-
ages only varied in regard to the semantic content of the visually depicted target
noun.

3.1.3. Self-paced reading procedure

In the noncumulative self-paced moving window task (Just, Carpenter, &
Woolley, 1982), participants read sentences one at a time in the center of a com-
puter screen with a series of dashes replacing all characters except for spaces. At
the start of each sentence, the participants saw a fixation cross (+) which upon
pressing the space bar disappeared to reveal the first word in the sentence. With
each consecutive press of the space bar the next word in the sentence appeared
and the prior word was replaced with dashes until the end of the sentence. Time
between button presses was recorded and reading times were calculated for each
word region. After each sentence the participants answered a comprehension
question that addressed the semantic content of the sentence. Participants com-
pleted six practice trials that were not included in analyses nor repeated as part of
the experimental stimuli.

3.1.4. Picture-selection task procedure

The picture-selection task was adapted from Morett and MacWhinney’s
looking-while-listening procedure. This task was designed to measure linguistic
predictive mechanisms.

At the start of each trial, the participants saw a 500 ms fixation cross (+),
which was replaced by a pair of images, one on the left and one on the right side
of the screen. The image presentations, respective to the screen, were vertically
centered and counterbalanced. Once the participant was familiar with the images,
they pressed the space bar to start the presentation of the auditory stimulus sentence delivered through individual headphones. The aural stimulus matched only one of the picture’s semantic content. The pictures remained on the screen until the participant selected one by pressing a specified key that corresponded to the side of the screen on which it appeared. Participants were instructed that they would see two images and needed to select “the picture that best matches the sentence they hear as quickly as possible” (see Figure 1). They were explicitly told that they did not need to wait till the end of the sentence to choose an image. In the experimental trials, the images either matched or mismatched in number of the critical noun. The participant’s button press was time-locked to the offset of the critical determiner in the auditory stimuli. Participants performed a short practice block of four sentences and data from the practice trials were not included in the analyses. Sentence comprehension was measured by participants’ picture selection accuracy.

Uninformative Condition

![Uninformative Condition Image]

Informative Condition

![Informative Condition Image]

Figure 1. Picture Selection task

3.1.5. Post-Screening Tests

After the sentence processing tasks the learners completed a final screening task - a grammar and vocabulary test to control for familiarity with the target words and grammatical constructions in the experimental sentences. Only learner participants who scored above 80% on the vocabulary section and 90% on the first grammar block were included in the study’s analyses leading to the exclusion of the data collected from 6 learners.
4. Results
4.1. Self-Paced Reading Task

To answer research question 1, results from the self-paced reading task were analyzed for both accuracy on the comprehension questions as well as for reading times in the critical N and N+1 regions.

4.1.1. Scoring and descriptives

Accuracy. After each stimulus sentence participants answered a yes-no comprehension questions which focused on the semantic content of different parts of the sentence (i.e., the verb, the subject, the adverbial phrase, and the prepositional phrase). One point was given for each correct answer with a total possible score of 32.

Reading times. The self-paced reading task produced reading times (RTs) for each word at the N-3, N-2, N-1, N, N+1, N+2, and N+3 with the critical noun region (N), and the preposition region (N+1) considered the critical regions of interest (see Example 5).

(5) Por la tarde el periodista…
... entrevistó a estos payaso*/s en el circo.

Analyses were run only those items for which participants gave correct responses to the comprehension questions, as it was important to assess processing during comprehension. The comprehension questions always focused on aspects of the sentence that did not include the number of the determiner or noun. RTs above 5000 ms and below 225 ms were excluded from analyses, which corresponded to 2.8% of the total RT values (the range is based on Blanchard, Rayner, & Pollatsek’s 1989 findings that single words require minimally 225 to 300 ms to be processed).

4.1.2. Inferential statistics

Accuracy. A two-way repeated measures ANOVA with a 2 (Agreement) x 2 (Construction type) was conducted to compare the effect of number agreement/violation and the type of linguistic construction (Demonstrative/Definite article) on comprehension question accuracy for each participant group. The findings revealed that there were neither significant main effects nor any interactions (p > .05), indicating that all participants were equally accurate in all conditions. This confirms that they were paying attention to the task and were able to understand the sentences and the questions. The type of linguistic construction or agreement had no effect on response accuracy.
Reading times. A 2 (Construction type) x 2 (Agreement/Violation) repeated measures ANOVA was run for each critical word position (N and N+1). Only analyses and findings of the N and N+1 region are discussed as they are the critical region of interest.

Natives. Significant main effects for Agreement were found in the N region \([F(1,31) = 8.48, p = .007]\) and in the N+1 region \([F(1,31) = 30.011, p < .001]\) revealing that the natives read significantly slower at N (Agreement: \(M = 460.42, SE = 20.00\); Violation: \(M = 501.96, SE = 28.11\)) and N+1 (Agreement: \(M = 424.76, SE = 15.7\); Violation: \(M = 528.128, SE = 23.88\)) when the number marking on the determiner did not agree with the noun (see Figure 2).

Advanced learners. There was a main effect for agreement at the N region \([F(1,18) = 5.130, p = .036]\) and the N+1 region \([F(1,18) = 4.962, p = .039]\) revealing that the advanced learners, like the natives, read significantly slower at N (Agreement: \(M = 625.64, SE = 60.13\); Violation: \(M = 732.911, SE = 83.91\)) and N+1 (Agreement: \(M = 484.75, SE = 37.89\); Violation: \(M = 565.96, SE = 666.35\)) when the number marking on the determiner did not agree with the noun (see Figure 3).

Intermediate learners. There was a main effect for agreement \([F(1,42) = 10.937, p = .002]\) and for construction type at the N region \([F(1,42) = 5.834, p = .02]\) but no main effects were found for either agreement \([F(1,42) = 3.777, p = .059]\) or for construction type \([F(1,42) = 1.167, p = .286]\) at the N+1 region. These findings reveal that intermediate learners, like the natives and the advanced learners, read significantly slower at N when the number marking on the determiner did not agree with the noun (Agreement: \(M = 653.06, SE = 31.03\); Violation: \(M = 736.44, SE = 36.73\)). Unlike the other groups, the intermediate learners also read significantly slower at N when the determiner was a demonstrative than when it was a definite article (Demonstrative: \(M = 719.02, SE = 35.72\); Definite Article: \(M = 670.48, SE = 30.33\)) (see Figure 4).
“entrevistó a los payaso*/s en el circo”
interviewed [D.O.M] the clown*/s in the circus.

Figure 2. SPR - RTs – Natives

“entrevistó a los payaso*/s en el circo”
interviewed [D.O.M] the clown*/s in the circus.

Figure 3. SPR – RTs – Advanced
4.2. Picture Selection Task

The results from the picture selection task were analyzed to answer research question 3 regarding whether intermediate and advanced L2 learners of Spanish are able to use number information from the demonstrative and the definite article to anticipate an upcoming target object to improve accuracy and response time of picture disambiguation.

4.2.1. Scoring and descriptives

Accuracy. Participants received one point for each correct picture selected for a total possible score of 32. All participants were able to match the aural stimuli to the visual images in over 75% of the cases.

Response times. Response times (RTs) were calculated as the time elapsed between the manually placed time-stamp at the offset of each critical determiner to the button press for each picture selection. Only RTs from correct trials were included in analysis. RTs below 225 ms were excluded from analyses, because they were likely to represent random selections initiated prior to the possible influence of the determiner (Haith, Wentworth, & Canfield, 1993; Matin, Shao, & Boff, 1993), which corresponded to 1.2% of the total RT values. The upper limit was determined by applying the outlier labeling rule (Hoaglin & Iglewicz, 1987) with the demarcation criteria set at $g = 2.2$ (see table 2).
4.2.2. Inferential statistics

**Accuracy**

**Native group.** A two-way repeated measures ANOVA with a 2 (Informativity) x 2 (Construction type) was conducted to compare the effect of cue informativity (informative / uninformative) and the type of linguistic construction (Definite Article/Demonstrative) on picture selection accuracy. The findings revealed that there were neither main effects nor any significant interaction \( (p > .05) \) indicating that native participants were equally accurate in all conditions.

**Advanced.** The advanced learners were more accurate in the when the cue was informative (Definite Article: \( M = 7.95, SD = .229 \); Demonstrative: \( M = 7.89, SD = .612 \)) than when it was uninformative (Definite Article: \( M = 7.79, SD = .535 \); Demonstrative: \( M = 7.53, SD = .315 \)) in both linguistic constructions.

**Intermediate learners.** Like the advanced learners, the intermediate learners were more accurate in the when the cue was informative (Definite Article: \( M = 7.49, SD = .768 \); Demonstrative: \( M = 7.42, SD = .763 \)) than when it was uninformative (Definite Article: \( M = 7.14, SD = .889 \); Demonstrative: \( M = 6.98, SD = 1.205 \)) in both linguistic constructions.

**Response times**

**Natives.** Significant main effects for Construction Type \( [F(1, 31) = 4.729, p = .037] \) and for Informativity \( [F(1, 31) = 7.948, p = .008] \) were found revealing that the natives responded faster in the demonstrative condition \( (M = 1048.61, SD = 384.1) \) than in the definite article condition \( (M = 1135.29, SD = 405.8) \) and in the informative condition \( (M = 1132.07, SD = 73.077) \) than in the uninformative condition \( (M = 1051.83, SD = 58.07) \).

**Advanced Learners.** A significant main effect for Informativity \( [F(1, 18) = 14.151, p < .001] \) was found revealing that the advanced learners, like the natives, responded faster in the informative condition \( (M = 1298.67, SE = 91.481) \) than in the uninformative condition \( (M = 1531.54, SE = 111.467) \). However, unlike the natives, there was no main effect for Construction type \( [F(1, 18) = 3.180, p = .091] \) nor a significant interaction between and Informativity and Construction type \( [F(1, 18) = .077, p = .785] \) (see Figure 15).

**Intermediate.** In the Intermediate Learner group there were no significant main effects for Informativity \( [F(1, 42) = .130, p = .720] \), Construction Type \( [F(1, 42) = .678, p = .415] \), nor for the interaction between the variables \( [F(1, 42) = 1.247, p = .270] \). Unlike the natives and the advanced learners, the intermediate
learners did not perform faster when number marking was informative: participants responded similarly in the Informative and Uninformative condition, and in both construction type conditions.

5. Discussion

This study set out to investigate whether L2 processing difficulties are the result of inefficiencies in integrating (bottom-up processing) or anticipating (top-down processing) linguistic information and whether such effects can be characterized by individual capacity limitations and language knowledge.

5.1. Research Question 1

All native speakers were sensitive to violations and demonstrated a reading slow down at the critical (N) and spill over (N+1) regions. This expected finding established the baseline for comparison for the learner groups. Both the advanced group and the intermediate group were also sensitive to the violation of number agreement between the determiner and the noun. However, sensitivity for the intermediate group was limited to the critical region only. These findings confirm that all participants were successful at establishing a number agreement relationship between the noun and its dependents inside the noun phrase. When this agreement relationship was not borne out due to a number mismatch, all participants were sensitive to this violation and demonstrated a reading slow-down in the critical and/or spill-over regions.

Contrary to the L1 effects found in Tokowicz and MacWhinney (2005) and Tokowicz and Warren (2010), all learners were equally sensitive to number violations in both the definite article and demonstrative condition. In summary, the major findings are in line with previous studies that have revealed that L2 learners can successfully detect morphosyntactic violations in real time when a certain level of proficiency has been reached (e.g., Alarcón, 2011; Keating, 2009; Osterhout et al., 2008; Rossi, 2006; Sabourin, 2003; Sabourin, Stowe, & De Haan, 2006; Sagarra & Herschensohn, 2011).

5.2. Research Question 2

The observed results revealed that the native group was significantly faster to select the target image in the Informative condition than in the Uninformative condition. This confirms that the native group exploited the number marking on the determiners to compute forward agreement relations between the determiner and the noun to narrow down the set of potential nouns that could follow and thus facilitate comprehension and correct picture selection. Their performance set the baseline for comparison. The advanced learners patterned very similarly to the natives confirming that they, too, were able to generate expectations about the upcoming noun from the morphosyntactic number cue on the determiner. In contrast, the intermediate learners did not demonstrate any response time advantage
in the Informative condition. This finding suggests that the intermediate learners have not yet developed the ability to utilize morphosyntactic number markings in an anticipatory manner. Turning to the analyses of L1 effects in picture-selection task, no effect of construction type was found for the learners. This finding supports the notion that the linguistic constructions were mostly too similar for L1 effects to be evidenced at the intermediate or advanced proficiency level.

In sum, L2 learners can successfully employ predictive mechanisms for the processing of morphosyntax in the L2. The strong claim of the RAGE hypothesis, that learners are systematically incapable of generating native-like expectations, is not supported by the findings of this study as it does not account for the specific processing situations (i.e., advanced proficiency) in which learners successfully employed anticipation.

5.3. Theoretical Framework

Returning to the broader question of why evidence of L2 anticipatory processing has not been robustly found in previous studies, I argue that although outside of the domain of morphosyntax there is some limited evidence that learners are capable of utilizing linguistic information, such as lexical-semantic information, in anticipatory processing (Hopp, 2015), the reason for which no such patterns have been found by learners within the domain of morphosyntax is attributable to previous studies’ narrow focus on gender, verbal aspect, and case-marking (e.g., Grüter, Lew-Williams, & Fernald, 2012; Hopp, 2015; Lew-Williams & Fernald, 2010). These structures contain morphosyntactic features notoriously difficult for L2 learners to acquire and automatize at even very advanced proficiencies, especially when they are absent or differ greatly from L1 features.

This current study has identified a linguistic phenomenon with which L2 learners were successful in exploiting morphosyntactic features to anticipate linguistic information. Both natives and advanced learners were able to exploit the number markings on the determiner - a morphosyntactic construction which is more easily acquired by learners than gender, case marking, or verbal aspect, due to some overlapping L1/L2 features - to accelerate disambiguation in the picture-selection task. The fact that the intermediate learners were not successful, combined with the observation that they were native-like in their detection of morphosyntactic violations in the self-paced reading task, supports the hypothesis that cognitive resources are dedicated first to integrative mechanisms. This is in line with the premise of the RAGE hypothesis, that resources exhausted in integrative processes impair or severely limit proactive processes such as prediction.

In summary, this study has disentangled two important processing mechanisms, integration and anticipation, for the first time in L2 studies and has found strong evidence to suggest that that integration processes are recruited prior to anticipatory mechanisms. It appears that the underlying causes of L2 divergent processing are external to the architecture of the L2 processor and can be localized in the interplay between proficiency and language experience.


