



*The Reading Matrix: An International Online Journal*  
Volume 17, Number 1, April 2017

## **Reading in Spanish as a Second Language: An Eye-Tracking Study**

**Elizabeth Enkin**

University of Nebraska-Lincoln

**Janet Nicol**

University of Arizona

**Zachary Brooks**

University of Arizona

**Kaitlyn Leigh Zavaleta**

University of Arizona South

### **ABSTRACT**

*In the present study, we examine sentence reading in low-proficiency Spanish learners using an eye-tracking methodology. This method reveals the real-time, uninterrupted process of reading comprehension, and can therefore shed light on L2 learners' functional proficiency. We created sentence pairs that were identical except for one word. The contrasting words in the sentence pairs differed in processing difficulty; in this case, the more difficult word had a lower frequency of occurrence. The effect of frequency on reading typically manifests as longer reading times, particularly when readers first encounter the critical word. The relative difficulty created by a low-frequency word is usually localized to the word itself. Results show a significant difficulty effect for both the learner group and control group. However, they also show that although the effect was localized in the control group, it was not localized in the learners: difficulty "spilled over" into the subsequent region, suggesting that these readers read at too fast a pace, which could lead to extensive re-reading. We believe that this pattern of reading could be remediated and suggest ways for doing so.*

### **INTRODUCTION**

In this paper, we examine language comprehension in adult novice learners of a second language (L2). Specifically, we examine sentence reading by relatively low-proficiency language learners and a control group of highly proficient language users. Exploring how well novice learners read written text is important because even the lowest levels of language instruction incorporate the written word, and beginning textbooks and online assignments may include reading passages that highlight certain vocabulary or some aspect of culture. In addition, assessments are likely to require students to read test questions and narratives. Yet virtually

nothing is known about how fluidly novice learners read their way through a sentence in a second language or in what way the requirement to read could impact their learning. The primary aim of the present research is to explore reading as it occurs, by using an eye-tracking methodology to examine eye movements during reading.

The process of reading in a second language is an under-studied area of research on second language processing, and the reading research that has been conducted tends to fall in one of two areas of study. One area explores the reading of complex sentences. The studies in this area are typically focused on sentence-comprehension strategies, strategies that guide the way incoming words are integrated into the ongoing structural representation of the sentence (e.g., Clahsen & Felser, 2006; Dallas, DeDe, & Nicol, 2013; Dussias & Cramer Scaltz, 2008; Dussias & Piñar, 2010; Felser, Cunnings, Batterham, & Clahsen, 2012; Felser & Roberts, 2007; Jiang, 2004; Juffs, 1998; Marinis, Roberts, Felser, & Clahsen, 2005; Pliatsikas & Marinis, 2013; Williams, 2006; Williams, Möbius, & Kim, 2001; Witzel, Witzel, & Nicol, 2012). The other area has focused on single-word reading, with the aim of developing models of the bilingual lexicon and word recognition processes in second language learners and bilinguals (e.g., Altenberg & Cairns, 1983; Dijkstra, Timmermans, & Schriefers, 2000; Dijkstra & van Heuven, 2002; Dijkstra, van Heuven, & Grainger, 1998; Duyck, Vanderelst, Desmet, & Hartsuiker, 2008; Forster & Jiang, 2001; Gollan, Forster, & Frost, 1997; Jiang & Forster, 2001; Kroll & Stewart, 1994; Mägiste, 1979; van Heuven, Dijkstra, & Grainger, 1998; von Studnitz & Green, 2002).

In the present study, we bridge these two areas by examining the speed and accuracy with which second language learners read words in sentences. Our main goal is to characterize the reading process in a second language in people who are experienced readers of their native language but novice readers in their second language. In order to do this, we created minimal-pair sentences that differed only in the difficulty of one of the words. Our aim was to look at the consequences of this difficulty on reading behavior.

We chose as a source of difficulty relative frequency of occurrence, for two reasons. One is that word frequency has such a reliable effect on processing time: frequently-occurring words are recognized more quickly than infrequently-occurring words. For example, although they are the same length and are similar in meaning, *cup* is a higher-frequency word than *mug*, and on average would be recognized faster. This effect is observable in lexical decision tasks, naming tasks, and reading tasks. In lexical decision tasks, participants determine whether a string of letters is a real word or not, in naming tasks, participants say aloud a printed word or the name of a pictured item, and in reading tasks, participants' reading times are measured (Frenck-Mestre, 2005; Gollan, Slattery, Goldenberg, van Assche, Duyck, & Rayner, 2011; Inhoff & Rayner, 1986; Lehtonen, Hultén, Rodríguez-Fornells, Cunillera, Tuomainen, & Laine, 2012; Libben & Titone, 2009; Murray & Forster, 2004; Paribakht, 2004; Schilling, Rayner, & Chumbley, 1998; Whitford & Titone, 2012).

The other reason is that although frequency effects are well documented in L2, reports differ as to whether there is a greater effect in L2 than the first language (L1), which is what is predicted by some models of lexical processing. A study by Van Wijnendaele and Brysbaert (2002) was among the first to examine the frequency effect (FE) in second language learners. In that study, Dutch-French and French-Dutch bilinguals were tested in a word-naming task, and larger frequency effects were found for L2 word naming; specifically, the difference in naming latencies for high-frequency words compared to low-frequency words was greater when participants named words in L2. Similar results are reported by Gollan, Montoya, Cera, and Sandoval (2008), for picture naming, and Duyck et al. (2008), for lexical decision. Duyck et al.

argue that this result is consistent with the idea that the FE is best described as a “logarithmic learning function”: Hence, a frequency difference at the higher end of the frequency range will have less impact than one at the lower end. Because L2 words are less frequent than L1 words—so the argument goes—there is a larger FE in L2. This would predict that novice learners, whose exposure to L2 words is very limited, should show very large frequency effects.

It should be pointed out that the finding of larger frequency effects in L2 may be tied, in part, to task demands, such as making a decision about lexical status or saying words out loud. The other elements of these tasks—beyond word recognition—could be difficult for second language learners, and this difficulty could interact with word frequency.

What, then, are the findings for tasks in which participants simply read for comprehension? There have been few studies that have examined the frequency effect of words embedded in L2 sentences, and the results vary. Gollan et al. (2011) tested high- or low-frequency words embedded in sentence contexts. The results showed that the magnitude of the frequency effect did not differ significantly across their participant groups despite differences in language proficiency. (There were, however, group differences in a production task with the same pairs of words, and in a lexical decision task.)

In contrast, an eye-tracking study by Whitford and Titone (2012) found a larger frequency effect in L2 than L1. Their study differed from the Gollan et al. (2011) research in a number of ways. Whitford and Titone tested an unusually large number of participants (125) and tested them in both L1 and L2. They also examined reading times for an unusually large number of items (about 100). Here is what they did: They created two sets of French and English paragraphs (in both sets, one was a translation of the other), and any given participant read one paragraph in English and one in French. Most of the content words in these paragraphs were coded for frequency. A central aim of the study was to compare English-French bilinguals—differing in degree of exposure to L2—reading in their L1 and L2. Therefore, the English and French versions of the paragraphs were equated on a number of dimensions (mean word frequency and length, degree of predictability). However, the high- and low-frequency words *within* each paragraph were not equated on dimensions such as predictability and plausibility, both of which could affect patterns of reading. So while their results are quite compelling—the same participants show a larger FE in L2 than in L1, and this difference is mediated by the extent of their exposure to L2—factors besides frequency of occurrence could be playing a role. Using a similar methodology and design, Cop, Keuleers, Drieghe, and Duyck (2015) likewise found that their bilinguals showed a significantly greater FE in L2 than in L1, though L2 proficiency had no significant effect on the size of the FE.

Overall, then, the findings for word-reading in context are mixed with respect to differences in the magnitude of the frequency effect in L1 vs. L2, and our study may shed further light on this issue.

### **Eye-tracking Studies of Sentence Reading**

In eye-tracking studies, eye movements are recorded as people read and comprehend sentences. Typically, readers fixate on a word (or words) for several hundred milliseconds (ms.) before they launch a saccade (“eye jump”) to the next word. Fixation duration varies as a function of word difficulty: words that are relatively infrequent are fixated for longer periods of time, as are words that are not predictable from prior sentence context. Information about one or more words may be registered during a fixation. Longer words usually require more than one

fixation. During a fixation, the fixated word is identified, and simultaneously, visual information about the upcoming word (the word to the right of the fixated word, in languages like English) is seen *parafoveally*. Visual resolution in the parafovea is poorer than in the fovea; however, some information about word form is typically registered, enough to allow readers to skip some words entirely. Skipped words are usually short and highly frequent (e.g. words like “the”). Not all reading is forward-going; it is estimated that about 10-15% of eye movements are regressive (Rayner, 1998). Difficulty of text matters. As difficulty increases, fixation durations become longer, the number of fixations (both forward and backward) increases, and the number of skipped words decreases (Juhasz & Rayner, 2006).

For the purpose of data analysis, sentences are divided into regions of interest (ROI), which are comprised of words or phrases that are then analyzed by various measures (Rayner, 1998). Effects of word frequency have been reported for *gaze duration* and *total time in a region*. *Gaze duration* is the sum of all fixations within a region before the reader moves to the next region. It is considered an “early” measure because it captures a reader’s first encounter with a word within a sentence. *Total time in a region* is the sum of fixations in a region before moving out of that region, plus the duration of fixations during re-reading. Because re-fixation durations are included in this measure, this is considered a “late measure”. We will consider both of these measures in our analyses. Another early measure that has been reported to reflect word frequency is *word-skipping*. As mentioned above, short, frequent words may be skipped entirely. We do not consider word-skipping here because our critical region contains both a noun and (gender-matched) determiner (see below for details), and readers are less likely to skip a phrase than a single word. Instead, we will examine the *number of fixations* within the critical region. If participants skip *either* the determiner or noun, this will be captured by *number of fixations*.

## Present Study

The main goal of this study is to investigate sentence reading in L2, focusing on the very much under-studied novice learners. We manipulate word frequency in order to create relative difficulty that has demonstrable effects on eye movements during reading. We examine data from three measures: gaze duration, total time in a region, and number of fixations. We expect L2 learners to read more slowly overall, and we may also see differences in patterns of reading. In addition, we expect that if frequency effects reflect a learning function that is logarithmic in nature, we should observe a larger frequency effect in the L2 learners.

## METHOD

### Participants

There were 26 participants (17 women, 9 men, mean age = 23.47). Participants were either affiliated with the University of Arizona or acquaintances of other subjects. There were two groups of participants: the Low-Proficiency group and the High-Proficiency group. All participants were volunteers and received a small monetary award at the end of the session, and they all had normal or corrected-to-normal vision.

The Low-Proficiency group consisted of 16 beginning-level learners of Spanish (native English speakers), and were recruited from introductory Spanish classes. Of these participants,

12 were enrolled in the second course within the Spanish course sequence (*Spanish 102*), 2 participants were enrolled in the first course of the sequence (*Spanish 101*), and 2 participants were not currently enrolled in a Spanish course but had previously had 2 years and 3 years of high school Spanish, respectively. Participants were late learners of Spanish, who began studying the language in school (between the ages of 10-16, mean age = 12.50). The second group was a control group of 10 High-Proficiency participants, consisting of 7 native Spanish speakers (NS) and 3 highly proficient Spanish speakers. They were recruited via email sent to Spanish department teaching assistants (TAs).

## Materials and Design

**Questionnaire.** A language background questionnaire was created to probe participants' demographic information, language history, and current language usage.

**Sentence-reading task.** Forty-eight word pairs were created in which the members of each pair differed in frequency. The pairs were synonyms or words that were similar enough in meaning to fit readily into the same sentence frame. The average frequency (per million) was 5148 for the higher-frequency items, and 1303 for the lower-frequency items (according to the Spanish corpus created by Mark Davies (2002) from Brigham Young University, which can be found at <http://www.corpusdelespanol.org>). In general, higher-frequency words are shorter, and shorter words elicit shorter reading times. So in order to ensure that any effects of frequency were not due to differences in length, we searched for words that were roughly equal in length; this was difficult given the other criteria that needed to be met. Ultimately, the average length of higher-frequency words was 5.98 letters and the average of lower-frequency words was 4.92 letters. Word pairs were selected based on consultation with graduate TAs who were teaching the Spanish 102 course (second semester Spanish), and reference to the textbook used for the elementary levels (*¡Dímelo tú!* by Nogales, Samaniego, and Blommers, 2008, 6<sup>th</sup> edition). Typically, second language learners at the beginning level know words ranging from simple items that denote color, food and restaurant items, clothing items, body parts, weather, to feelings, descriptors for a basic place such as a house, and everyday words (such as “question” and “answer”), to more complex (less frequent) words such as those used when needing to describe a situation in detail, such as an emergency.

The words were embedded in sentence contexts that varied with respect to structure. Following standard counter-balancing procedures, each sentence appeared with the higher-frequency word (of a word pair) in one presentation list, and with the lower-frequency word in the other presentation list (Table 1 below contains an example). Each list contained 24 higher-frequency and 24 lower-frequency words. In addition to the 48 experimental items, there were 24 “filler” sentences (which were designed to explore a different research question) and 4 practice sentences (to familiarize participants with the task). To ensure that participants attended to the sentences, 18 of the 72 sentences were followed by a YES/NO comprehension question. Responses to the comprehension questions were not analyzed as these questions were only intended to keep participants paying attention.

**Table 1.** Sample Sentence with Higher- and Lower-frequency Items

HIGHER-FREQUENCY ITEM	LOWER-FREQUENCY ITEM
Normalmente al chico le gusta la <u>lluvia</u> , pero hoy tiene frío.	Normalmente al chico le gusta la <u>nieve</u> , pero hoy tiene frío.

---

*Usually the boy likes the rain, but today  $\phi$   
(he) is cold.*

*Usually the boy likes the snow, but today  $\phi$   
(he) is cold.*

---

## Procedure

**Questionnaire.** The questionnaire was completed by participants before the sentence-reading task.

**Sentence-reading task.** Each sentence was presented with standard capitalization and punctuation. The sentences appeared as single lines of text on a 21-inch CRT monitor. Participants were directed to read each sentence silently at their natural reading speed. They were told to make sure that they could comprehend each sentence well enough so that they would be able to accurately answer an occasional YES/NO question. Participants' eye movements were recorded from the right eye using a Dr. Bouis Monocular Oculometer at a sampling rate of 200 Hz. There was an approximate distance of 60 cm from the participants' eye to the monitor, which allowed for single-character resolution. In order to minimize head movements, a bite plate and head rest were used. At the beginning of each session, the eye-tracker was calibrated for the participant, and then recalibrated after every four trials. Each trial began with a cue to fixate on the far left edge of the display. This fixation mark was intended to alert participants of the location of the beginning of the sentence. After the presentation of the fixation mark, an entire sentence was then displayed. The first character of the sentence appeared one space to the right of the fixation point. As soon as they had finished reading each sentence, participants pressed a button on a button-box. The sentence then disappeared, and was replaced either by a string of dashes or a YES/NO comprehension question (to which participants responded via the press of a button on a button-box), followed by feedback about whether they had answered correctly.

## RESULTS

For purposes of analysis, sentences were divided into regions; the critical word (always a noun) appeared with a determiner in Region 2 (sentences contained either 4 or 5 regions). As described earlier, we consider two "early" measures, Gaze Duration (the sum of fixation durations in a region before moving out of the region), and Number of Fixations in a region before moving out of the region; and a "later" measure, Total Time in Region (gaze duration, plus the duration of re-reading fixations). Means for these measures for the two groups appear in Table 2.

**Table 2.** Mean Reading Times for Gaze Duration and Total Time in Region for the Critical Region and Mean Number of Fixations in the Critical Region

Group	Word Type	Gaze Duration (ms. per word)	Number of Fixations (per region)	Total Time in Region (ms.)
High-Proficiency	Higher-Freq.	194	1.78	554
	Lower-Freq.	228	2.11	667
	<i>Frequency Effect</i>	34	.33	113

<b>Low-Proficiency</b>	Higher-Freq.	325	2.42	930
	Lower-Freq.	373	2.59	1021
	Frequency	48	.21	91
	Effect			

Analyses of variance were conducted for each measure for each group. Each ANOVA was conducted with participants as the random factor (“*F1*”) and with items (“*F2*”) as the random factor. Analyses for each group will be addressed in turn.

**High-Proficiency Group:** All comparisons show an effect of frequency. ANOVAs reveal significant frequency effects for Gaze Duration ( $F1(1,9) = 12.971, p < .006; F2(1,47) = 6.73, p < .013$ ), Total Time in Region ( $F1(1,9) = 19.52, p < .002; F2(1,47) = 15.385, p < .001$ ), and Number of Fixations ( $F1(1,9) = 11.336, p < .008; F2(1,47) = 6.73, p < .013$ ).

**Low-Proficiency Group:** This group also showed a frequency effect on all three measures. For Gaze Duration,  $F1(1,15) = 9.152, p < .009; F2(1,47) = 5.018, p < .03$ , for Total Time in Region,  $F1(1,15) = 8.383, p < .011; F2(1,47) = 4.855, p < .033$ , and for Number of Fixations,  $F1(1,15) = 9.266, p < .008; F2(1,47) = 5.66, p < .021$ .

Group effects were also analyzed. For Gaze Duration (the only measure that shows a greater frequency effect for the Low-Proficiency group), the critical interaction of Group x Frequency was not significant ( $p$ 's  $> .4$ ), although there was a main effect of both Group ( $F1(1,24) = 13.79, p < .002; F2(1,47) = 157.7, p < .0001$ ) and Frequency ( $F1(1,24) = 16.897, p < .0001; F2(1,47) = 7.668, p < .008$ ).

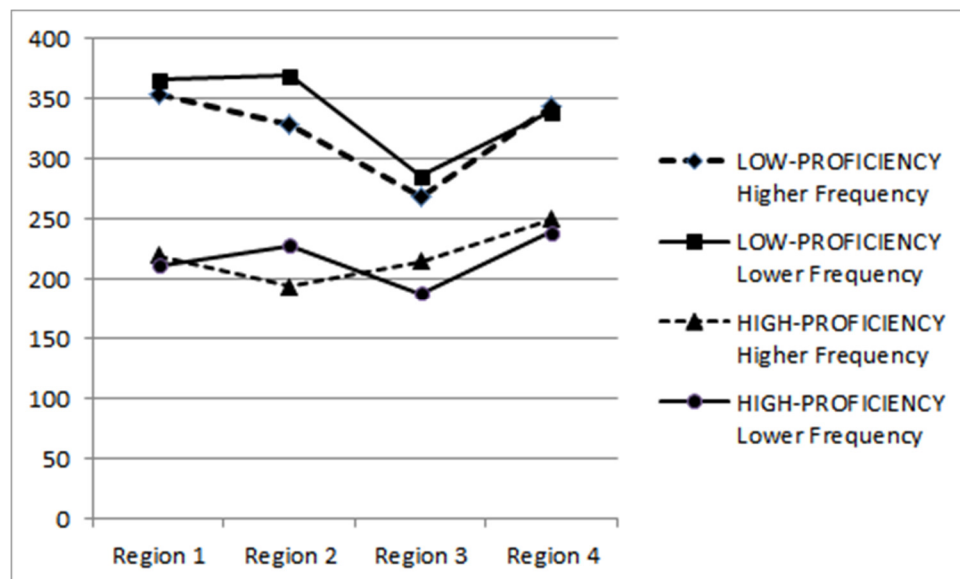
The other two measures show the same pattern. For Number of Fixations, the main effects of Group and Frequency were significant (for Group,  $F1(1,24) = 10.68, p < .004; F2(1,47) = 53.25, p < .0001$ , and for Frequency,  $F1(1,24) = 15.53, p < .002; F2(1,47) = 8.89, p < .006$ ), but the interaction was not ( $p$ 's  $> .26$ ). For Total Time in Region, there was a main effect of Group ( $F1(1,24) = 50.49, p < .0001; F2(1,47) = 192.88, p < .0001$ ) and Frequency ( $F1(1,24) = 16.44, p < .0001; F2(1,47) = 10.02, p < .004$ ). The interaction was nonsignificant ( $p$ 's  $> .6$ ).

In order to explore the broader impact of word frequency on sentence reading, we charted gaze duration times for the first four sentence regions. An example sentence illustrating the four regions appears in Table 3. Figure 1 shows the means for the two proficiency groups.

**Table 3.** Example Sentence, Region by Region

Region 1	Region 2	Region 3	Region 4	(Region 5)
Normalmente al chico le gusta	la lluvia/la nieve,	pero	hoy tiene	(frío.)

**Figure 1.** Gaze Duration (ms. per word) for Regions 1-4. The frequency manipulation appeared in Region 2.



As reported earlier, both groups showed a significant frequency effect; this is reflected in Region 2. But the groups show different effects in the subsequent region, Region 3: the Low-Proiciency group shows a residual effect of frequency in this region, but the High-Proiciency group shows no such effect. In fact, for the High-Proiciency group, the frequency effect appears to be reversed (this reversal is not statistically significant). The difference in the frequency effect for these groups in Region 3 (that is, the interaction between Group and Frequency) is statistically significant:  $F(1,24) = 6.4, p = .018$ ;  $F(1,47) = 6.194, p = .016$ . In the fourth region, the frequency effect is diminished or absent for both groups.

## DISCUSSION

One goal of this study was to explore text reading in language learners by examining whether the FE in low-proficiency speakers is manifest in the same way as in high-proficiency speakers. Our results suggest that it is, and that it is not.

In keeping with prior studies, we analyzed the effect of frequency on the initial and later encounters with the critical words. We considered two “early” measures and one “late” measure of reading comprehension. Our results show a FE for both low- and high-proficiency groups for all measures, with no indication of a greater effect for the low-proficiency learners. This is consistent with eye-tracking research by Gollan et al. (2011), despite several major differences between the studies, including the generally higher L2 proficiency of their low-proficiency participants (who reported their “percentage daily reading” in L2 at 13.9%, while L2 reading for



our participants was largely restricted to the reading they did in the classroom), and vastly lower word frequencies (the average of their high-frequency words was about 80 occurrences per million; the average for low-frequency words was about 10) in their test materials.

However, our results are not consistent with the eye-tracking results reported by Whitford and Titone (2012). Recall, their study used naturalistic paragraphs in which words were coded for frequency. While their English and French paragraphs were equated for mean frequency of most of the content words, the sentence contexts in which the high-frequency vs. low-frequency words appeared were not. If lower-frequency words happened to be less predictable, or less plausible, or both, they could cause relatively greater processing difficulty for the L2 reader on those grounds alone. In other words, perhaps it is not the FE that is larger in L2, but rather the effect of predictability or plausibility. Likewise, these factors may have also played a role in Cop et al.'s (2015) eye-tracking study's results, which were similar to those of Whitford and Titone's: bilinguals showed a larger FE in L2.

Where our two groups differed was in the processing of material in the subsequent region: the low-proficiency group shows further processing slowdown due to word frequency; the high-proficiency group does not. Let us consider the performance of each group in turn. The low-proficiency group shows what is known as a "spillover effect". "*Spillover*" refers to the finding that the difficulty incurred by one word "spills over" into subsequent regions. Typically, the effect of frequency of occurrence is localized on the word itself. A lower-frequency word takes longer to recognize but once recognized, the item is no longer difficult and the reader moves the eyes rightward. But this is not always the case. Within the "monolingual" literature, the findings for proficient adult readers appear to be inconsistent, with some studies reporting such effects and others reporting no effect. For example, Rayner, Sereno, Morris, Schmauder, and Clifton (1989), Kennison and Clifton (1995), and Pollatsek, Juhasz, Reichle, Machacek, and Rayner (2008) showed a FE spilling over from an adjective to the adjacent rightward noun, and Henderson and Ferreira (1990) reported frequency-related spillover from noun to verb. Kliegl, Nuthmann, and Engbert's (2006) multiple regressions analysis of reading times on most of the words in each of 144 German sentences (in the Potsdam Sentence Corpus) shows spillover effects related to frequency, as well as factors such as predictability. But other studies have shown limited effects (e.g. Williams & Morris, 2004), or no spillover effects at all (e.g. Henderson & Ferreira, 1993; Raney & Rayner, 1995; Slattery, Pollatsek, & Rayner, 2007).

One study showed group differences in the frequency-spillover effect. Ashby, Rayner, and Clifton (2005) compared the performance of "highly skilled" vs. "average" readers. Grouping was based on scores on the Nelson-Denny Reading Test, a standardized test of reading comprehension in high school and college students. Similar to our study, they embedded pairs of high- and low-frequency words into the same sentence contexts, but they also manipulated the predictability of the words. In the condition in which a word was not predictable from context, frequency affected the skilled and average readers differently: the skilled readers showed a FE on the word itself but no spillover effect; the average readers showed no FE on the word itself, but a significant spillover effect. The authors suggest that "the average readers spent relatively less time looking at low-frequency unpredictable words and often recognized them only after moving to the next word" (p. 1075). In other words, the average readers did not wait for lexical recognition processes to finish before launching a saccade to the next word.

Similarly, although our low-proficiency participants show a FE on the word itself, they also tend to move on before word recognition is completed. Our high-proficiency group, on the other hand, does not move ahead prematurely. In fact, it is likely that they begin to process

material in the *subsequent region* while they are still fixating on the critical word. Recall that the high-proficiency group shows a “cross-over” effect: they show significantly *shorter* gaze durations during the reading of higher-frequency words (vs. lower-frequency words), but *longer* gaze durations in the following region. One reason for this could be that while they are fixated on the critical word, they begin to process the upcoming word. If the current word is lower in frequency, it may require more than one fixation, and the second fixation would be closer to the next word, allowing for greater preview. But why wouldn’t the low-proficiency group show the same benefit? After all, this group also showed more fixations on lower-frequency words. But preview benefits are also correlated with “foveal load” (Henderson & Ferreira, 1990). The greater the processing difficulty of the current word, the less beneficial the “parafoveal” preview. The low-proficiency group should, in general, suffer greater “foveal load” and thereby benefit less from preview.

In sum, the reading patterns of the two groups appear to be as follows. The high-proficiency group is more likely to allow lexical recognition processes to run to completion before launching a saccade to the next word. Lower-frequency words will require longer fixations and sometimes more than one. Sometimes the additional fixation will land them closer to the next word, and they may begin to process that word in parallel. In contrast, the low-proficiency group does not always wait for word recognition to complete before moving to the next word. If word recognition is still in progress when the eye moves to the next word, parafoveal processing of the next word is unlikely.

However, if this scenario is right, does this not mean that the frequency effect is actually larger for the low-proficiency group than it appeared? Perhaps if we consider the FE for the critical and subsequent regions together (added, not averaged) the low-proficiency group will show a significantly larger effect than the control group. We conducted this analysis. Specifically, we compared the combined FE (in the critical plus spillover regions) for the low-proficiency group to the FE in only the critical region for the high-proficiency group. Analyses of variance showed no significant difference, either on the by-subjects analysis or the by-participants analysis ( $p$ 's > .16). But note the following: even if the L2 FE had been significantly larger than the L1 FE, one would need to interpret this with caution. Typically, slower processing (in terms of reading or responding to a stimulus) is associated with larger effects. This issue has been discussed in some detail by Chapman, Chapman, Curran, and Miller (1994).

Why do L2 readers show a spillover effect? We can only speculate about this, but perhaps they are used to reading at a certain pace, and despite the difficulty of reading in L2, do not increase early fixation durations beyond a certain point. One consequence of this would be spillover, but another would be significantly more time spent re-reading, just as we observed. This aligns with reports that in general, L2 readers engage in significantly more re-reading than native speakers (see, e.g., Frenck-Mestre, 2002), and excessive regressive eye movements to earlier parts of a sentence is not an efficient way to comprehend text.

A pedagogical implication of these results is that language instructors may want to focus on aiding L2 readers fully process each word before advancing to the next one. Since some online (real-time) reading methodologies that are used for psycholinguistic experiments do not allow looking-ahead or backtracking in a sentence, they might in fact be useful for L2 teaching. One such task is the “self-paced moving-window display” (Just, Carpenter, & Woolley, 1982), where readers advance through a sentence word-by-word at their own pace, without being able to look-ahead or backtrack. This encourages readers to finish processing before moving on. Another possible task could be the “maze task” (Forster, Guerrero, & Elliot, 2009), which is an

alternative to the self-paced reading procedure. In this task, sentences are also presented incrementally word-by-word, and readers must complete the processing of the current word before progressing forward. However, in this task, after reading the first word in a sentence readers advance to the next screen where they see two words presented horizontally alongside each other. Here, they must choose the correct word out of the two that is the natural grammatical continuation of the sentence (the other word would be ungrammatical and unnatural). This procedure continues until the end of the sentence has been reached.<sup>1</sup> Although the maze task is not a highly natural reading technique, it may offer an advantage over self-paced reading insofar as ensuring no spillover effects. That is, since this task requires readers to process sentences word-by-word by making incremental grammatical decisions, each word must be fully integrated with the previous words in the sentence before moving on. The maze task has the added advantage that it is more engaging than self-paced reading and can be used as a teaching tool (Enkin, 2016; Enkin & Forster, 2014).

## CONCLUSION

The main goal of this research was to determine whether the reading patterns of low-proficiency second language learners resemble those of highly proficient readers. We targeted the FE because this is such a robust effect in the visual word recognition and reading literature. Our results indicate that both groups show a FE, of similar magnitude. But the timing of the effect differs, with high-proficiency readers showing a localized effect on the word itself, and low-proficiency readers showing an effect across two regions: the critical word and the subsequent region. This might suggest that L2 readers advance forward through text before they are ready, which could lead to extensive re-reading. Foreign language teachers may therefore want to consider the results of this research when planning reading-based lessons and the time allotted for them. Furthermore, potentially useful teaching interventions might be the self-paced moving-window display and the maze task, which do not allow look-ahead or backtracking, thereby encouraging readers to finish processing each word before moving on.

*Elizabeth Enkin is an Assistant Professor of Spanish Applied Linguistics in the Department of Modern Languages and Literatures at the University of Nebraska-Lincoln. Her research is on foreign language learning and teaching. Her research projects have focused on novice-level L2 sentence processing, advanced-level online L2 learning, and TA professional development.*

Email: [eenkin@unl.edu](mailto:eenkin@unl.edu)

*Janet Nicol is a professor at the University of Arizona, in the departments of Linguistics and Psychology, the Program in Cognitive Science, and the Interdisciplinary Program in Second Language Acquisition and Teaching. Her research examines L2 acquisition and processing, using methods such as elicited production, eye-tracking, and Event-Related*

---

<sup>1</sup> A demonstration of the maze task can be found at the following web link:  
[www.u.arizona.edu/~kforster/MAZE](http://www.u.arizona.edu/~kforster/MAZE)

*Potentials.*

Email: [nicol@u.arizona.edu](mailto:nicol@u.arizona.edu)

**Zachary Brooks** has a PhD in Second Language Acquisition with PhD minors in Cognitive Science and Linguistics (University of Arizona). His dissertation brings together the fields of applied linguistics, cognitive science, and management.

Email: [zbrooks@email.arizona.edu](mailto:zbrooks@email.arizona.edu)

**Kaitlyn Leigh Zavaleta**, PhD in Psychology (University of Arizona), is an adjunct professor at several universities, including the University of Arizona's South campus in the Psychology department. She researches on the process of L2 and L3 acquisition, specifically as it is altered by individual differences in cognition and language experience.

Email: [kdalton@email.arizona.edu](mailto:kdalton@email.arizona.edu)

## ACKNOWLEDGEMENTS

We would like to thank Alexis Kreun, Bonnie Holmes, Jose Herrera, Lara Petersen, and Kyle Krone for their help on this project. We would also like to thank the University of Arizona Social and Behavioral Sciences Research Institute (SBSRI) for a Pre-doctoral Research Grant, which allowed us to compensate subjects for their participation.

## REFERENCES

- Altenberg, E. P., & Cairns, H. S. (1983). The effects of phonotactic constraints on lexical processing in bilingual and monolingual subjects. *Journal of Verbal Learning and Verbal Behavior*, 22(2), 174-188.
- Ashby, J., Rayner, K., & Clifton, C., Jr. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 58(6), 1065-1086.
- Chapman, L. J., Chapman, J. P., Curran, T. E., & Miller, M. B. (1994). Do children and the elderly show heightened semantic priming? How to answer the question. *Developmental Review*, 14(2), 159-185.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27(1), 3-42.
- Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, 22(5), 1216-1234.
- Dallas, A., DeDe, G., & Nicol, J. (2013). An event-related potential (ERP) investigation of filler-gap processing in native and second language speakers. *Language Learning*, 63(4), 766-799.

- Davies, M. (2002). Corpus del español: 100 million words, 1200s-1900s. <http://www.corpusdelespanol.org>.
- Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*, 42(4), 445-464.
- Dijkstra T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175-197.
- Dijkstra, T., van Heuven, W. J. B., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*, 38(3-4), 177-196.
- Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, 128(3), 501-513.
- Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of *wh*-gaps by Chinese-English second language speakers. *Second Language Research*, 26(4), 443-472.
- Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, 15(4), 850-855.
- Enkin, E. (2016). Second language learning with the story maze task: The training effect of weaving through stories. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée*, 19(1), 1-21.
- Enkin, E., & Forster, K. (2014). The maze task: Examining the training effect of using a psycholinguistic experimental technique for second language learning. *Journal of Linguistics and Language Teaching*, 5(2), 161-180.
- Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, 34(1), 67-98.
- Felser, C., & Roberts, L. (2007). Processing *wh*-dependencies in a second language: A cross-modal priming study. *Second Language Research*, 23(1), 9-36.
- Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, 41(1), 163-171.
- Forster, K. I., & Jiang, N. (2001). The nature of the bilingual lexicon: Experiments with the masked priming paradigm. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 72-83). Oxford: Blackwell.
- Frenc-Mestre, C. (2002). An on-line look at sentence processing in the second language. *Advances in Psychology*, 134, 217-236.
- Frenc-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research*, 21(2), 175-198.
- Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1122-1139.
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58(3), 787-814.

- Gollan, T. H., Slattery, T. J., Goldenberg, D., van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, *140*(2), 186-209.
- Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 417-429.
- Henderson, J. M., & Ferreira, F. (1993). Eye movement control during reading: Fixation measures reflect foveal but not parafoveal processing difficulty. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *47*(2), 201-221.
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*(6), 431-439.
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(4), 603-634.
- Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language*, *44*(1), 32-51.
- Juffs, A. (1998). Some effects of first language argument structure and morphosyntax on second language sentence processing. *Second Language Research*, *14*(4), 406-424.
- Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, *13*(7-8), 846-863.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228-238.
- Kennison, S. M., & Clifton, C., Jr. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 68-81.
- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12-35.
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149-174.
- Lehtonen, M., Hultén, A., Rodríguez-Fornells, A., Cunillera, T., Tuomainen, J., & Laine, M. (2012). Differences in word recognition between early bilinguals and monolinguals: Behavioral and ERP evidence. *Neuropsychologia*, *50*(7), 1362-1371.
- Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 381-390.
- Mägiste, E. (1979). The competing language systems of the multilingual: A developmental study of decoding and encoding processes. *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 79-89.
- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language processing. *Studies in Second Language Acquisition*, *27*(1), 53-78.
- Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, *111*(3), 721-756.
- Nogales, F. R., Samaniego, F. A., & Blommers, T. J. (2008). *¡Dímelo tú!* Boston, MA: Heinle Cengage Learning.

- Paribakht, T. S. (2004). The role of grammar in second language lexical processing. *RELC Journal*, 35(2), 149-160.
- 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.
- Pollatsek, A., Juhasz, B. J., Reichle, E., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: A reversed delayed effect of word length. *Journal of Experimental Psychology: Human Perception and Performance*, 34(3), 726-750.
- Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 49(2), 151-172.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C., Jr. (1989). Eye movements and on-line language comprehension processes [Special Issue]. *Language and Cognitive Processes*, 4(3-4), 21-49.
- Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, 26(6), 1270-1281.
- Slattery, T. J., Pollatsek, A., & Rayner, K. (2007). The effect of the frequencies of three consecutive content words on eye movements during reading. *Memory & Cognition*, 35(6), 1283-1292.
- van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39(3), 458-483.
- Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance*, 28(3), 616-627.
- von Studnitz, R. E., & Green, D. W. (2002). The cost of switching language in a semantic categorization task. *Bilingualism: Language and Cognition*, 5(3), 241-251.
- Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, 19(1), 73-80.
- Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, 9(1), 71-88.
- Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English *wh*-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, 22(4), 509-540.
- Williams, R., & Morris, R. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, 16(1-2), 312-339.
- Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics*, 33(2), 419-456.