Second Language Learning with the Story Maze Task: Examining the Training Effect of Weaving Through Stories

Elizabeth Enkin
University of Nebraska-Lincoln

Abstract

The maze task is a psycholinguistic experimental procedure that measures real-time incremental sentence processing. The task has recently been tested as a language learning tool with promising results. Therefore, the present study examines the merits of a contextualized version of this task: the story maze. The findings are consistent with previous research results (Enkin & Forster, 2014), and highlight the merits of using the contextualized maze version for language (Spanish) learning. Specifically, a story maze training-test paradigm revealed that learners trained on structures differing from their native language (English) showed little difference in reaction times between similar-to-English and different-from-English structures as compared to learners who were trained on structures similar to English, thus showing that the story maze task may help students learn constructions that may pose processing difficulty. Quantitative and qualitative survey data further showed that learners found the task highly engaging, thereby emphasizing its promising usefulness.

Résumé

La tâche des labyrinthes est une procédure expérimentale en psycholinguistique qui mesure en temps réel la compréhension croissante de phrases. La tâche a récemment été testée comme outil d'apprentissage d'une langue seconde et offre des résultats prometteurs. Par conséquent, la présente étude examine les mérites d'une version contextualisée de cette tâche : le labyrinthe d'histoires. Les résultats sont compatibles avec les conclusions d'une étude antérieure (Enkin et Forster, 2014) et mettent en lumière les mérites de l'utilisation de la version contextualisée des labyrinthes pour l'apprentissage d'une langue étrangère (l'espagnol). Spécifiquement, un paradigme de test formation de labyrinthe d'histoires a révélé que les apprenants formés à des structures différentes de leur langue maternelle (l'anglais) ont montré peu de différence en temps de réaction entre des structures similaires à l'anglais ou différentes de l'anglais, comparé à des apprenants formés à des structures similaires à l'anglais. Ces résultats montrent ainsi que la tâche des labyrinthes d'histoires peut aider les étudiants à apprendre des constructions posant des difficultés de compréhension. Par ailleurs, des résultats de recherche quantitative et qualitative ont montré que les apprenants trouvaient cette tâche très captivante, soulignant ainsi son utilité prometteuse.

Second Language Learning with the Story Maze Task: Examining the Training Effect of Weaving Through Stories

Introduction

The focus of this article is on how a task, called the story maze, can be used to assist with second language (L2) learning. The story maze task (described in further detail later) is a modified version of a maze task, which is a procedure that is used in psycholinguistic experiments to measure reaction times (in milliseconds) as subjects read and comprehend sentences (Forster, 2010; Forster, Guerrera, & Elliot, 2009; J. Witzel & Forster, 2014; N. Witzel, Witzel, & Forster, 2012; Qiao, Shen, & Forster, 2012). This task is called an online task because it aims to measure real-time processing of sentences (as opposed to offline tasks where subjects have more time to reflect on their responses—see Marinis, 2003, for a review of online tasks). In the maze task, participants are asked to weave their way through a sentence by selecting the correct grammatical alternative from two choices, thereby presenting a "maze" to participants. As is typical for psycholinguistic tasks that aim to measure real-time processing speed, subjects are asked to complete the maze task as quickly as possible, though not so quickly that a mistake is made.

The task visually presents two words alongside each other, and participants must then decide, by pressing arrow keys, which word out of the two could correctly continue the sentence. One of the choices is the correct alternative, whereas the other word would be ungrammatical as well as unnatural when taking into consideration the prior context (i.e., the words that have already been chosen for the sentence). It should be noted that this version of the maze task is called the grammaticality maze (G-maze), where the incorrect word alternative is ungrammatical, as opposed to the lexicality maze (L-maze) where the incorrect choice is a legal nonword (N. Witzel et al., 2012). Once a correct response is given, these two words disappear and are replaced by two new words, and this procedure continues until the participant has reached the end of the sentence. Figure 1 below illustrates an example item.

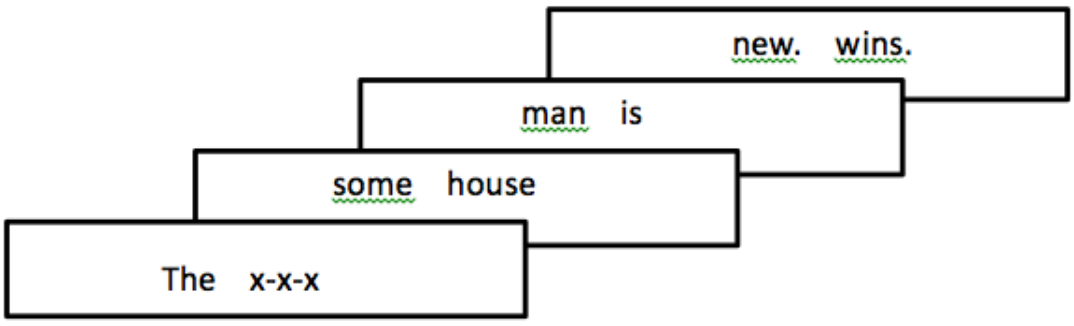

Figure 1. A sample maze task sentence, frame by frame.

In each frame, only one word can correctly continue the sentence. Participants view each frame separately and can only view the next frame once a correct choice has been selected. Here the sentence would be: *The house is new.*

In Figure 1, participants would first see a frame where the first word (in this case *The*) would be displayed on the left side of the screen, and alongside it *x-x-x* would appear. This indicates to the participants that this is the first word in the sentence, and that they are free to press any key to continue. Immediately after that, the subsequent frame, which contains two words side by side, replaces the previous frame (in this case, the frame would be *some* and *house*). Participants must now choose which word can continue the sentence (i.e., *house*). This procedure goes on until the end of the sentence is reached (here, the sentence would read *The house is new.*). When used in experimental settings, if participants erroneously choose an incorrect alternative, the trial is aborted, and participants will automatically be directed to begin the next sentence (a live demonstration can be found at the following website: http://www.u.arizona.edu/~kforster/MAZE/).

In psycholinguistic experiments, most sentence processing paradigms must make use of comprehension questions in order to ensure that participants actually comprehend what they are reading (e.g., eye tracking [Rayner, 1998]; self-paced reading [Just, Carpenter, & Woolley, 1982]). However, due to its incremental nature, the maze task does not need to use them. That is, comprehension questions are considered unnecessary since learners realize that as they process each individual word, they must also process the meaning of that word. This is because, in the next frame, they will be asked to continue building that sentence (Forster et al., 2009). Indeed, the maze task has yielded results also shown by multiple other methodologies (that utilize comprehension questions): Forster et al. (2009) reported that the maze task showed garden path effects corresponding to those found through eye tracking. Nicol, Forster, and Veres (1997) used the maze task to show that structures leading to errors in production (i.e., subject-verb agreement errors that are mostly caused by a mismatch in which singular head nouns are followed by plural nonhead nouns, rather than when plural head nouns are followed by singular nonhead nouns) can also affect processing time in comprehension. In their study, relative to the control conditions where there were no mismatches, processing time for correct responses to verbs in sentences increased only when a singular head noun was followed by a plural nonhead noun, thereby supporting results from production research. N. Witzel et al. (2012) found that the maze task, eye tracking, and self-paced reading are all effective in detecting processing difficulties in temporarily ambiguous structures, but the maze task has the advantage of producing highly robust localized effects. This latter finding further underscores that the maze task facilitates an environment where comprehension must occur at each word in a sentence.

Since the maze task requires incremental, deep comprehension of sentences, my previous research has shown that the task can indeed be used successfully for language learning purposes (Enkin & Forster, 2014; and in an earlier working draft of that paper, Enkin, 2012). By training beginner learners of Spanish for a period of time with the maze task, benefits were seen on several measures (namely, a post-training maze task session, an untimed grammaticality judgment task, and a paper-and-pencil based pretest-posttest fill-in-the-blank task). Learners were very welcoming of the task, which led us to argue that the maze task should be considered as a potential language learning task. We suggest that further research should be carried out where a more complete, video game-like program is created and tested. In an effort to continue this line of research, the present study is

therefore devoted to investigating how a modified version of the maze task, the story maze, can be used for language learning purposes.

The Maze Task and Language Training

When the maze task is used for language practice (or training), it requires learners to incrementally select the correct continuation of a sentence, from two choices, as quickly and as accurately as possible. However, unlike using the task purely as a technique for measuring reaction times, when participants erroneously choose the incorrect alternative, they are asked to try the sentence again rather than having the trial aborted. My previous research has argued that it is through this method that the maze task training can provide essential practice that may reinforce, or complement, formal instruction (Enkin & Forster, 2014). In other words, practice through the maze task may be able to strengthen the connections between existing associations made during class instruction. Therefore, the task could be a potentially powerful tool when used as outside practice for structures that are being taught in class. Moreover, in addition to invoking comprehension processes, the maze task may also invoke production processes: learners must make incremental grammatical decisions, between two word choices, as they are constructing a sentence, thereby requiring a selection process, a mechanism that is necessary during production (see also Nicol et al., 1997, for discussion about the maze task possibly requiring production mechanisms). Thus, the task may have broad benefits for L2 learning.

In my co-authored maze task study discussed above (Enkin & Forster, 2014), we hypothesized that because the maze task asks for rapid responses, it could be used as a training instrument to develop language automaticity and fluency in learners. Although fluent language use is an important goal for L2 learners, relatively few training studies have been conducted in this area (Akamatsu, 2008; De Jong & Perfetti, 2011; DeKeyser, 1997; Robinson, 1997). The above maze task study (i.e., Enkin & Forster, 2014) used a maze training-test paradigm to train second-semester Spanish learners on specific structures with the maze task. A training effect was found through an analysis of the test session, where reaction times to trained and untrained structures were compared (the structures and training effect are elaborated upon further below when describing the hypothesis for the present study). We therefore reasoned that because the task asks for rapid responses, it could serve not only as a training tool, but also as a post-training test of implicit knowledge (see Ellis, 2009, and Rebuschat, 2013, for summaries of tasks that are considered measurements of implicit knowledge). Implicit knowledge (the intuitive knowledge of how a language works) is unlike explicit knowledge (the knowledge of grammatical rules), and is generally thought to be the necessary knowledge base that is drawn upon during fluent language use (Ellis, 1993, 2005). The claim that the maze task training can build implicit knowledge and automaticity is re-examined in the Discussion section.

The Present Study: The Story Maze Task

The story maze task is a contextualized version of the maze task, which has been created keeping the language learner in mind. That is, although it offers the same advantages of rapid and incremental processing as the maze task does, it has the additional advantage of involving the reader in a contextualized environment. This environment further facilitates processing sentences for meaning, which has been shown to be an

effective instructional technique for language learning (VanPatten, 2004; VanPatten & Cadierno, 1993; VanPatten, Collopy, Price, Borst, & Qualin, 2013). There are three changes that I have made to the maze task in order to create the story maze task. First, the task presents participants with logically sequenced sentences that make up a story (four sentences comprise a story). Second, it incorporates a picture at the beginning of each sentence with the rationale that this further facilitates processing of the stimuli, since research has shown that pictures may aid in the L2 learning process, particularly for vocabulary learning (Theriault, 2009). Fletcher and Tobias (2005) have also discussed the advantage of learning through both words and pictures as compared to with words only, and Gambrell and Jawitz (1993) showed that pictures can be used effectively in comprehending stories. I also thought that including pictures would add to the overall experience of the task given previous research that has found that students enjoyed learning vocabulary with the use of pictures (Bohinski, 2012). Lastly, for the third change, if participants have trouble constructing any given sentence, the task allows learners to view that sentence in full on the screen. In this way, learners are never removed from the contextualized environment (further details of the task are presented in the Method section).

Because this version of the maze task is contextualized, but still requires immediate responses in a rather plain format, it lends itself well to the mobile phone application (app) arena, where a video game-like environment can be created. Thus, creating and testing the story maze task makes an important contribution to the field of L2 learning since the benefits of creating an app for learners may be related to increasing motivation for language practice. Indeed, as Tremblay and Gardner (1995) suggested, instructors may want to focus on pedagogical instruments that enhance language learning motivation.

University-aged learners are very enthusiastic to use technology when possible, and they are more willing and eager, and therefore motivated, to complete tasks outside the classroom when there is technology involved (Blake, 2012). In fact, research on text-based synchronous chat has shown that students showed increased levels of involvement in these environments (e.g. Abrams, 2001; Blake, 2000; Smith, 2003). As technology has evolved, so has the array of technological tools, including video games, which are now used for learning and are very applicable for language learning (Gee, 2007; Sykes & Reinhardt, 2013). Thus, this study is quite timely considering the continual development of new technological tools that serve as motivational instruments for language learners.

Because the maze procedure requires rapid responses (participants are time pressured and any time limit can be set for all choices [i.e., each frame]), it is a type of language practice that is usually only available through interaction. However, learners (especially beginners) may be too anxious to seek out language practice that involves speaking and listening. Therefore, although the story maze task is not meant to substitute for this critically important human interaction, it may be an activity that can be done online to supplement it.

Given the potential benefit of this task for language learners, and specifically for college-aged beginners, the central question in the present study is how effective a story maze task can be for language learning with beginner (second-semester) university-aged learners of Spanish. By using the training-test design from the Enkin and Forster (2014) study, reaction times on the post-training story maze test session are measured. In the Enkin and Forster study the constructions of interest in the training-test design had two different versions (similar-to-English and different-from-English—these are discussed below), and during training, participants were divided into two groups and were trained on one of these

versions. A test session including both versions of the structures was then given to all participants. However, training sessions were not counter-balanced insofar as the content of the sentences for the two groups was not kept as similar as possible. Counter-balancing would have been optimal in order to better assess the effect of each training version on the test session. In the present study, training stories are counter-balanced for both training groups (stimuli discussed below). Additionally, this study includes more experimental items than the Enkin and Forster study, thus allowing for more data to be collected. Moreover, using the story maze task rather than the maze task was hypothesized to be more engaging for learners due to its contextualized nature.

Measuring the Training Effect: Training-Test Paradigm Sentence Types and Hypothesis

The stimuli for the training-test paradigm used in measuring the story maze task training effect were created taking into account the hypothesis that the maze task measures rapid and intuitive responses. Evidence suggests that L2 learners may have difficulty storing different-from-English structures as procedural representations (Tokowicz & MacWhinney, 2005), and so I thought that these types of structures could be used in investigating a training effect. Participants were trained on either similar-to-English or different-from-English structures in a counter-balanced design. Interest was placed on the different-from-English structures since those were hypothesized to be problematic for learners. Based on this reasoning as well as on previous results (Enkin & Forster, 2014), my hypothesis was that on the test session, participants trained on different-from-English structures would find both types of structures comparable (as compared to those trained on similar-to-English structures), whereas participants trained on similar-to-English structures would find different-from-English structures more difficult. Below is an explanation of the four sentence constructions used. These were specifically chosen for the proficiency level of the students: All of the constructions had been previously covered in their course (Spanish 102), before the study began mid-semester. There was no additional class instruction focusing particularly on these structures after the study started.

Four different sentence constructions in Spanish were used: object relative clauses, direct object pronouns, copulative verbs *ser* and *estar*, and the prepositions *para* and *por* [*for*].[1] The purpose of using four different constructions was so that there would be variety within the stories since each story was composed of four sentences. There was no set order for constructions within stories (thereby avoiding predictability), but each story contained a sentence with each construction. Each of the four constructions had a similar-to-English version (henceforth, English-similar) and a different-from-English version (henceforth, Spanish-specific), which referred to the degree of similarity to the participants' native language (L1; English). I hypothesized that participants would show the expected training effect discussed above when analyzing each sentence construction separately. Table 1 contains example stories taken from training sessions, and illustrates the difference between English-similar and Spanish-specific versions.

Two of the four constructions, namely object relative clauses and direct object pronouns, can be structurally either similar-to-English or different-from-English. With the object relative clauses, the English-similar versions contained an overt subject after the relative pronoun, whereas the Spanish-specific versions contained a pro-drop situation (Spanish is a pro-drop language but English is not). This particular construction was used

because in an object relative clause, pro-drop may make the sentence appear, at first glance to a beginner, like a subject relative clause. Thus, beginners must learn that verb conjugation, especially in this context, is important for interpreting the sentence correctly. However, this is not the critical element to look for in English because in order to have an object relative clause in English, an overt subject (noun or pronoun) must be used after the relative pronoun. For direct object pronouns, the English-similar versions contained a clitic that appeared in the same location as the object pronoun would in an English sentence (object pronouns are postverbal in English). For the Spanish-specific versions, the direct object pronoun was raised (unlike in English, which does not allow this placement). Sentences containing direct object pronouns utilized an infinitive verb after a tensed verb because in this formation in Spanish, a clitic can appear either attached at the end of an infinitive verb (postverbal) or raised before the tensed verb (preverbal).

The other two constructions, copulative verbs and the prepositions *para* and *por* [*for*], are lexical contrasts between English and Spanish. In the case of the copulative verbs, *ser* and *estar*, these verbs both translate as *to be* in English. However, they have different meanings—*ser* expresses permanency while *estar* expresses temporary states (this is a distinction that English does not express with two forms of the copula). One of these verbs, *ser*, is more readily assimilated into its English translation, and moreover, as learners go through a series of developmental stages with acquiring *ser* and *estar*, those of English L1 background tend to linger at the stage where they exclusively use and overuse *ser* (VanPatten, 1987). As VanPatten (1987) suggested, one hypothesis is that, at this stage in acquisition, L1 influence may play a role because there would be an overlap of the L1 and L2. That is, at this stage of learning, only one copula exists (just like in English), which is why Spanish learners of English L1 background linger here. In light of this argument, *ser* is used in English-similar versions of the verb *to be*. Sentences focused on specific uses of the verb, such as describing occupations, expressing time, and defining relationships and personality traits. The Spanish-specific versions contained its counterpart, *estar*, and these sentences focused on uses such as describing emotions and feelings, expressing location, and using *estar* with gerunds.[2]

With regard to *para* and *por* [*for*], Mumin (2011) discussed how English L1 students learning Spanish face persistent problems when learning these two prepositions. As Mumin noted, one of the major reasons for this issue is L1 interference. That is, only one preposition, *for*, exists in English, which is equivalent to both *para* and *por* in Spanish. However, *para* and *por* can have different translations as well. In the participants' beginner Spanish class, there were many more definitions given for *por* (as opposed to *para*) that are not translated as *for* (though *for* is also included as a definition). For example, students learned that *por* can be translated as *for*, *in*, *by means of*, and *through*. Far less variation exists for *para*, which can be translated as *for* and *in order to*. Due to this difference, in the present study, the English-similar sentences utilized *para* used as *for*, whereas the Spanish-specific sentences focused on *por*, which was used in sentences as *for*, but also as *through*, *by means of*, *during*, *in*, and *by*.

Table 1
Example Stories (Sentence by Sentence) for Both Training Groups

Sentence Types	English-Similar Sentences	Spanish-Specific Sentences
To Be (*ser* vs. *estar*)	Mario y Verónica *son* novios y hoy tienen su boda grande. Mario and Veronica *are* fiancées and today they have their big wedding.	Mario y Verónica *están* muy contentos porque hoy tienen su boda grande. Mario and Veronica *are* very happy because today they have their big wedding.
Object Relative Clause (with vs. without subject pronoun)	El lugar de la boda *que ellos escogieron* tiene muchas flores. The place of the wedding *that they chose* has many flowers.	El lugar de la boda *que (Ø) escogieron* tiene muchas flores. The place of the wedding *that Ø (they) chose* has many flowers.
Preposition (*para* vs. *por*)	Muchas personas vienen *para* la recepción porque quieren celebrar el día especial. Many people come *for* the reception because they want to celebrate the special day.	Muchas personas pasan *por* la recepción porque quieren celebrar el día especial. Many people come *by* the reception because they want to celebrate the special day.
Direct Object Pronoun (pronoun that follows vs. precedes a verb)	Sus mejores amigos traen el champán y todos juntos *van a beberlo*. Their best friends bring the champagne and all together *they are going to drink it*.	Sus mejores amigos traen el champán y todos juntos *lo van a beber.* Their best friends bring the champagne and all together *they are going to drink it*.

Method

Participants

Twenty-three undergraduate students who were enrolled in one Spanish 102 class (a beginner level, second-semester Spanish class) at a large university in the United States Southwest participated for course credit. Participants were L1 speakers of English. They were randomly assigned into one of two training groups, which was either "English" or "Spanish" (11 students in the English training group and 12 students in the Spanish training group). These groups' names referred to the type of sentences that students received during story maze training sessions. The English training group received sentence structures that were English-similar, whereas the Spanish training group received Spanish-specific structures. Therefore, all participants received training, but were assigned to one of these two groups.

The Spanish 102 course followed a communicative teaching approach and was taught fully in Spanish. Grammar instruction included outlining grammatical rules to learners, and then the remainder of class time was primarily devoted to practice in context. Class activities helped develop basic language skills (i.e., listening, speaking, reading, and writing) and included both pair and small group work focusing on interaction.

Proficiency test. All students entering a Spanish course at this university, including Spanish 102, must take a 20-25 minute computer-adaptive placement (proficiency) exam. This web-based exam is the BYU (Brigham Young University) WebCAPE (Computer-Adaptive Placement Exam) and is administered by the university. The exam asks questions of varying difficulty levels and therefore adapts its questions according to students' responses. The qualifying score to be placed into Spanish 102 is a range of 201 to 309. Any score below a 201 places students into Spanish 101, scores of 310 to 479 place students in intermediate level courses, and scores above a 480 place students into third-year (advanced) classes. The only alternative way that students can enter into the 102 level is if they have college transfer credits from the previous level of Spanish (101). This proficiency test served to ensure that the students in this study were of comparable Spanish skill level.

Materials and Design

Story maze task training. Both training groups completed one training session a week for 3 weeks. There were a total of 24 sentences that comprised each training session (a total of six stories with four sentences each; 18 total sentences per construction across 3 weeks, or 72 total sentences across 3 weeks). Each training group received story maze training on their sentence types. That is, the English training group was trained on the English-similar types, whereas the Spanish training group was trained on the Spanish-specific types. However, the stories' sentences contained either the exact same or very similar content so as to keep the context of each sentence the same for both groups (this is illustrated in Table 1). Before the first word of each sentence was shown, a picture illustrating the content of the sentence was displayed to participants. The purpose of displaying pictures was to create an enhanced contextualized environment and to give more of an app-like feel to the task. Because the content of stories (and therefore each sentence) was the same (or very similar) for both groups, the same illustrations were used for both groups (e.g., in Table 1, the first sentence for both English-similar and Spanish-specific types was accompanied by the same picture of a happy bride and groom embracing).

Each of the three sessions contained the same stories for each group, and the incorrect alternatives were exactly the same for both groups, when possible. That is, sometimes, alternate versions of the same story required different incorrect word alternatives. When exactly the same incorrect alternatives were not possible, care was taken to only use distracter words that would appear at some point in both training groups. All incorrect alternatives were appropriate for the Spanish 102 level. Examples (a) and (b) below illustrate sentences from each training group, and include incorrect alternatives. In these examples, incorrect alternatives are the second word in each frame (frames are illustrated by " / "). During the training sessions, participants were given the option of trying the sentence again if they made a mistake. If a mistake were made, the program would stop and show an error message immediately, so that students could see where in the sentence they had made the mistake.

(a) *El x-x-x / lugar agresivas / de documentos / la y / boda unos / que medicina / ellos pero / escogieron rutina / tiene duérmete / muchas llegamos / flores. escuchamos.*

(b) *El x-x-x / lugar agresivas / de documentos / la y / boda unos / que medicina / escogieron pero / tiene duérmete / muchas llegamos / flores. escuchamos.*

Story maze task test. One post-training story maze task test session was administered to all participants on the fourth week after the 3-week training period. This session contained all new sentences (and therefore, new stories and pictures), and there was an equal amount of both English-similar and Spanish-specific types. There were a total of eight stories (32 sentences total), each with unique content, with four stories (16 sentences total) containing English-similar types and the other four stories (another 16 sentences total) containing Spanish-specific types. Again, incorrect alternatives were of appropriate level. For this session, however, participants were not able to try the sentences again if they made a mistake. When a mistake occurred, immediate feedback was provided in the form of an error message on the screen. As in the training sessions, this type of feedback indicated to the learners where in the sentence the error had occurred.

Procedure

Story maze task training. The story maze task training sessions (and the post-training test session) were run using the DMDX software package, which was developed by Forster and Forster (2003) at the University of Arizona. By using remote testing, each session was sent through email as a link. When participants clicked on the link, DMDX automatically installed on their PCs for the duration of the task. Participants completed each training session in one sitting and only one time, but they were given a full week to complete each session, thereby allowing them to complete each session at their convenience.

The items were presented in black letters on a white background. Each item displayed frame-by-frame made up a sentence of a four-sentence story. The first frame of each sentence in a story displayed a picture, which illustrated the content of the upcoming sentence (e.g., for the first sentence in Table 1, a picture of a happy bride and groom embracing would be shown). Participants pressed the right arrow key to proceed when they were ready, and the next frame showed the first word of the sentence with "x-x-x" alongside it. Participants pressed the right arrow key to move on to the subsequent frame when ready. After these first two frames, each following frame contained two words that were side by side: one was the correct next word in the sentence, whereas the other was both grammatically and semantically incorrect (see example below with English gloss: sentence is, "The wedding place that they chose has many flowers.").

El x-x-x / lugar agresivas / de documentos / la y / boda unos / que medicina / ellos pero / escogieron rutina / tiene duérmete / muchas llegamos / flores. escuchamos.

The x-x-x / place aggressive / of documents / the and / wedding a / that medicine /
they but / chose routine / has [go to sleep] / many [we arrive] / flowers. [we listen.]).

The correct and incorrect alternatives appeared randomly on either the left-hand side or on the right-hand side of the screen. Furthermore, since training sessions contained the same sentences and incorrect alternatives for each training group, the incorrect alternatives appeared on random sides (left or right) of the screen from session to session. Through this procedure, participants could not memorize the correct alternative's position on the screen. Stories were presented in a randomized order for each subject for each session (although the stories themselves were presented linearly, from sentence one to four).

After viewing the picture and first word, participants were instructed to choose the correct word in each frame that would logically continue the sentence, as quickly and as accurately as possible, by pushing the left or right arrow key (for either the word on the left-hand side of the screen or the right-hand side, respectively). If the word was correctly selected, the next frame (showing two new words) was displayed immediately. However, if the incorrect alternative was selected, an error message (i.e., " < < error > >") was displayed, and participants were then given the option of trying the sentence again by pushing the right arrow key. If participants decided to not try the sentence again, they pushed the left arrow key, and that entire sentence, in red, was displayed in full on the screen (e.g., "The sentence was: This was the sentence."). The rationale here was that this procedure would keep learners in the contextualized environment of the story. Participants were then instructed to press the right arrow key when they were ready to move on to the next item (i.e., a new picture indicating the beginning of a new sentence either in that story or in a new story). If participants made the correct choice throughout the frames for an item, the final frame was followed by a correct message (i.e., "CORRECT"). Subsequently, the beginning of the next item would appear. All participants received the same two practice stories (a total of eight sentences) at the beginning of each session. These practice items were of appropriate level, but did not include any of the experimental sentence structures (see below for example sentence and English gloss: sentence is, "Three friends want to start a new rock band.").

Tres x-x-x / amigas un / quieren clase / empezar pero / una sabemos / nueva sus /
banda levantarse / de tienes / rock. al.

Three x-x-x / friends a / want class / [to start] but / a [we know] / new your / band [to wake up] / of [you have] / rock. [to the].

Story maze task test. The link for the story maze test session (i.e., the post-training session) was also sent through email, and participants had one week during which they could complete this session (again in one sitting). All participants completed this and previous sessions (all sessions had to be completed in order to be included in the study). The instructions remained the same and items were presented in the same manner as in the training sessions. Once again, stories were presented in a randomized order for each subject. The only difference, however, was that in this session, participants were not given the choice of trying a sentence again. If an error occurred, the program would display an

error message like in the training sessions, which was immediately followed by the entire sentence displayed in red on the screen. As in the training sessions, participants would then be asked to press the right arrow key to move on to the next item. Participants were given two practice stories (a total of eight sentences) at the start of the session, and again, these were level appropriate and did not include the key structures.

Results and Discussion

Results: Story Maze Task Test Analysis

The analysis was carried out using linear mixed effects models (LMERs) that were fitted to the data points. This analysis was carried out using the LMER function in the lme4 package for the R program (Baayen, 2008a; Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000; R Core Team, 2013). The method using LMERs is applicable for this analysis since it offers an essential advantage over the traditional F1/F2 method, which is used with ANOVAs. Linear mixed effects models allow for two crossed random effects (subjects and items in this case) to both be treated as random effects within the same model. Furthermore, the LMER software will analyze the data for each individual trial and will not need to aggregate over items and subjects. The software will then arrive at the best fitting linear model, which will include both subjects and items as random effects. The *p* values for the effects were generated by using Markov Chain Monte Carlo (MCMC) simulation, which uses 10,000 iterations, from the languageR package (Baayen, 2008b; Baayen et al., 2008). In the analysis presented below, a model was first run that included an interaction term between the factors. If the interaction was significant, the interaction term was kept, and therefore, main effects were generated with a model that included the interaction. Otherwise, an analysis only including the main effects would need to be carried out.

Prior to the analysis, the raw reaction times (RTs) were log converted so as to correct for the marked positive skew, which is characteristic of reaction time data. The RTs shown were back transformed to the original scale for ease of interpretation. They were trimmed such that those times under 300 milliseconds (ms) and those over 5000 ms were not included in the analysis, which resulted in the removal of 0.67% of the data. All trials (i.e., frames) where an error occurred were then discarded, resulting in the removal of an additional 3.63% of the data, and trials that were never seen due to a prior error were all ignored. Participants would not see trials if they would "error out" of a sentence, and therefore would not continue constructing that sentence. Additionally, all frames that presented sentences in full to participants (this would happen after errors occurred) were not included in the analysis. The first word in each sentence (i.e., trials where the correct response was provided for the subject) as well as trials where subjects would view the accompanying pictures for each sentence were also removed from the analysis.

Linear mixed effects models were fitted to the data, with subjects and items both as random effects. Training group was a fixed-effect, between-subjects factor, with the levels English (for the training group that received English-similar sentences) and Spanish (for the training group that received Spanish-specific sentences). Sentence type was analyzed as a second fixed-effect, within-subjects factor, with the levels English-similar types and Spanish-specific types.

Raw reaction times to each word in each sentence that were responded to correctly were used as the dependent variable (since the primary focus was on the overall training

effect, rather than on individual sentence regions). The critical interaction of training group by sentence type was highly significant ($t = 6.11$, $p < .001$), which reflected a training effect (no main effect of sentence type, $t = 1.43$, $p > .05$, nor of training group, $t = 0.48$, $p > .05$). This result indicated that the difference in RTs between the two training groups for the Spanish-specific sentences (87 ms) was significantly greater than the difference for the English-similar sentences (7 ms). Stated more clearly, the interaction illustrated that there was little difference in difficulty (6 ms) between English-similar and Spanish-specific sentences for the Spanish training group, but that there was a substantial difference (88 ms) for the English training group. Figure 2 shows the back transformed log mean RTs. The means demonstrated that learners who were trained on Spanish-specific versions (i.e., participants in the Spanish training group) yielded comparable RTs for both Spanish-specific and English-similar versions. However, learners that received training on English-similar structures (i.e., the English training group) found the Spanish-specific structures more difficult. It should also be noted that the direction of the differences discussed above (6 ms and 88 ms) was not the same: The Spanish training group showed lower RTs for Spanish-specific types as compared to English-similar types; the reverse was true for the English training group.

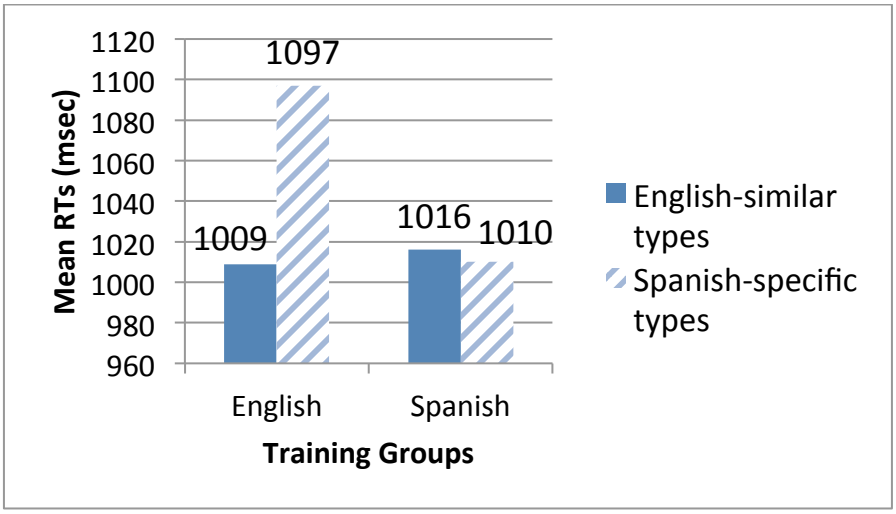

Figure 2. Mean reaction times (in ms) on English-similar and Spanish-specific sentence types for the story maze task post-training test session. English and Spanish training groups refer to the sentence types that participants received during story maze task training sessions (English-similar or Spanish-specific).

Story maze task test: Analysis by sentence construction. Using the same method of analysis described above, four separate analyses for each sentence construction revealed the same significant training group by sentence type interaction (no significant main effects—all $p > .05$): direct object pronoun ($t = 2.36$, $p < .05$; main effect of sentence type, $t = 0.46$, and training group, $t = 0.47$), object relative clause ($t = 2.65$, $p < .05$; main effect of sentence type, $t = 0.28$, and training group, $t = 0.60$), preposition ($t = 2.62$, $p < .01$; main effect of sentence type, $t = 1.64$, and training group, $t = 0.65$), and *to be* verb ($t = 4.68$, $p < .001$; main effect of sentence type, $t = 0.23$, and training group, $t = 0.19$). For each sentence construction, once again, the means indicated little difference in difficulty between English-

similar and Spanish-specific types for the Spanish training group, but a substantial difference for the English training group. Table 2 shows the back transformed log mean RTs and the differences between each sentence type: English-similar or Spanish-specific. The negative number differences indicate changes in direction such that the mean RTs were numerically lower for Spanish-specific types as compared to English-similar types.

Table 2

Mean RTs and Differences (in ms) for Each Construction for Both English-Similar and Spanish-Specific Sentence Types for Each Training Group (English and Spanish)

	English			Spanish		
	English-Similar	Spanish-Specific	Difference	English-Similar	Spanish-Specific	Difference
DOP	1008	1088	80	1033	1046	13
ORC	1042	1109	67	1044	1028	-16
Preposition	1012	1129	117	974	981	7
To be	974	1060	86	1020	984	-36

Note. DOP = Direct object pronoun, ORC = Object relative clause.

Discussion

Echoing previous results showing the maze task is useful for L2 learning, the present findings show that the story maze task helps learners process L2 sentences quicker, specifically those with structures unlike the L1 (Spanish-specific). One explanation is that since these structures may be difficult to store as procedural representations, they may require practice, but learners may already know English-similar structures, and further practice on them will not enable faster reaction times. Interestingly, four separate analyses for each construction also show the same overall results, indicating the story maze task's usefulness in learning each construction.

Although this task measures rapid responses, the results may not indicate gains in implicit knowledge or automaticity, as suggested in Enkin and Forster (2014), given that much repetition and time are needed to form implicit knowledge (see Ellis, 2002, for discussion). Participants in this study were only trained for 3 weeks on 18 sentences per construction. Moreover, since the maze requires incremental linguistic decisions to be made at the conscious level, and thus does not mimic the natural reading process (Forster et al., 2009; Witzel & Forster, 2014), maze training may not actually form implicit knowledge.

Reaction Questionnaire

Given the maze task's unnaturalness, I thought that students might perhaps find the story maze task uninteresting. Thus, participants completed a questionnaire during class time about the story maze task. Fourteen questions measured feedback on the likeability, usefulness, and overall perceptions of the task. The questions required ratings (*5* to *1*: *5 = strong yes*, *4 = yes*, *3 = neutral*, *2 = no*, *1 = strong no*). All questions appear in the Appendix, but overall, these questions asked participants:

- if they found the task fun and helpful for learning Spanish;
- if they thought the task was comparable to, or better than, more traditional online homework;
- if they liked pictures and storylines and if they thought they help with sentence construction;
- if they enjoyed completing the task every week;
- if they thought the task could help on performance elsewhere (such as on exams); and
- whether the task could be part of a Spanish curriculum and/or be useful for other languages.

One open-ended question asked participants to write any other comments they wanted to share.

The questionnaire yielded an average score of 4.6 out of 5 on all questions. Top scoring questions revealed that students found the story maze task was an enjoyable supplement to online workbooks (which may be due to its contextualized video game nature), that they found the task fun and helpful, and that they thought the pictures and storylines were fun and helped with completing the task. Learners also indicated that they thought the task could be used successfully in a Spanish curriculum, and that the task may be helpful for learning other languages. With respect to the open-ended responses, no student included a negative comment about the task. Table 3 provides direct quotes from learners.

Table 3
Questionnaire Student Comments

"It was engaging and made me focus."
"It was fun!"
"Helpful program!"
"It is more interesting as you go through sessions."
"I love this so much more than online workbook."
"The graphics helped me understand the stories."
"Helps to read sentences aloud so that you hear them and that makes you go faster."
"I couldn't get enough of it!"
"It was AWESOME!"

Despite (or perhaps because of) the task's unnatural nature, it was still highly enjoyable and engaging: Students became excited about their own learning. From the learner's perspective, it could be that this task provides a break from routine-scheduled assignments, but it could also be that students enjoy the video game nature of the task. With comments such as "it is more interesting as you go through sessions," learners seem to be "getting into" the activity, thereby creating conditions where they learn through gaming.

Pedagogical Implications and Future Research

Both the training effect and receptive attitude found for the story maze task illustrate that the task may have a place in the L2 classroom. These results are quite timely

considering the need for video game-like innovations in the field of L2 learning and instruction (Thorne, Black, & Sykes, 2009). To add to homework variation, the story maze task could provide learners with a different type of controlled language practice (that they enjoy) in addition to more traditional online workbooks. This may be a promising option, and is highlighted by one student comment: "I love this so much more than [the] online workbook." Moreover, because the task asks for rapid responses, and because learners seem to think of it as a game (taking as evidence comments such as "It is more interesting as you go through sessions" and "I couldn't get enough of it!"), the task's future may be a mobile phone app. Indeed, the story maze task is a more engaging version of the maze task that lends itself well to the app arena, which is evidenced not only by the high rating on the questionnaire, but also from comments such as "The graphics helped me understand the stories." Furthermore, if educators become interested in the product, it may be helpful for online language learning, specifically with respect to hybrid language courses where some portion of in-class hours are substituted by online work and activities (see, e.g., Blake, 2011, 2012, and Rubio & Thoms, 2012, for discussions on hybrid classes).

The story maze task can also be an important stepping-stone for learners with respect to what tasks they are able to perform in the L2. In a study by Toyoda and Harrison (2002) where Japanese learners participated in a gaming environment involving interaction through texts, one important finding was that due to learners' limited language proficiency, some had trouble keeping up with the game. As opposed to more sophisticated language video games (see Peterson, 2010, for a review of games), the story maze task provides the best of two worlds. That is, it is not an entire video game by itself; it is a game-like task that learners are motivated to complete during their transitional period between traditional workbook types of activities and more interactive full-fledged language games. It may therefore be useful for language classes, especially for more beginner and intermediate levels.

Future research is important. One question is: Can training effects be seen elsewhere, such as on production tasks? This could be examined through a pretest-posttest design using a control group. Indeed, as discussed earlier, production processes may play a role in the task, especially given one open-ended comment from a participant: "Helps to read sentences aloud so that you hear them and that makes you go faster." On a related note, because the task may provide students with interaction-type practice as discussed earlier as well, another project could be to create an auditory maze, which may further help learners develop both listening and speaking skills.

Conclusion

This study's results suggest that practice through the story maze task can be very enjoyable for students, and results in learning benefits. This research is therefore not only encouraging, but also very timely, considering the developing research area of language learning and technology, and the benefits that learning through technology can have for students. In fact, in a study about hybrid introductory Spanish language courses, Scida and Saury (2006) found that after students completed outside-of-class online activities focusing on grammar and vocabulary, they were more prepared for class and were more confident in their abilities, which created a more effective and enjoyable class experience. The authors concluded that it would be worthwhile to include web-based activities in all language courses (and not only in hybrid settings). Thus, in the future, a preliminary mobile phone

app of the story maze task could be created for the purpose of supplementary out-of-class practice, which could then serve as a platform for development.

Correspondence should be addressed to Elizabeth Enkin.
Email: eenkin@unl.edu

Notes

[1] The last sentence construction was the only one not used in Enkin and Forster (2014).

[2] *Ser* and *estar* can both be followed by adjectives, but *ser* takes an adjective that denotes properties that are permanent, whereas *estar* takes an adjective that denotes properties that are more temporary. Incorrect alternatives were therefore never adjectives, after either verb. Instead, a conjugated verb (though never a gerund) was chosen as the incorrect alternative after either verb. In the case of *estar* followed by a gerund (i.e., a verb), a noun was the incorrect alternative (since a noun cannot follow *estar*).

References

Abrams, Z. (2001). Computer-mediated communication and group journals: Expanding the repertoire of participant roles. *System, 29*(4), 489-503.

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics, 29*(2), 175-193.

Baayen, R. H. (2008a). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, United Kingdom: Cambridge University Press.

Baayen, R. H. (2008b). *languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics"*. R package version 0.953.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Blake, R. (2000). Computer-mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology, 4*(1), 120-136.

Blake, R. (2011). Current trends in online language learning. *Annual Review of Applied Linguistics, 31*, 19-35.

Blake, R. (2012). Best practices in online learning: Is it for everyone? In F. Rubio & J. J. Thoms (Eds.), *Hybrid language teaching and learning: Exploring theoretical, pedagogical and curricular issues* (pp. 10-26). Boston, MA: Heinle Cengage Learning.

Bohinski, C. A. (2012). *The portrait of a word: The use of mental and visual images in the acquisition of form, meaning, and use of Spanish concrete nouns* (Unpublished doctoral dissertation). Temple University, Philadelphia, PA.

De Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning, 61*(2), 533-568.

DeKeyser, R. (1997) Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*(2), 195-221.

Ellis, R. (1993). A theory of instructed second language acquisition. In N. C. Ellis (Ed.), *Implicit and explicit language learning* (pp. 79-114). San Diego, CA: Academic Press.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition, 24*(2)*,* 223-236.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*(2), 141-172.

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing, and teaching* (pp. 3-25). Tonawanda, NY: Multilingual Matters.

Enkin, E. (2012). The maze task: Training methods for second language learning. *Arizona Working Papers in Second Language Acquisition & Teaching*, *19*(5), 56-81.

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.

Fletcher, J. D., & Tobias, S. (2005). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 117-133). New York, NY: Cambridge University Press.

Forster, K. I. (2010). Using a maze task to track lexical and sentence processing. *The Mental Lexicon*, *5*(3), 347-357.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116-124.

Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods, 41*(1), 163-171.

Gambrell, L. B., & Jawitz, P. B. (1993). Mental imagery, text illustrations, and children's story comprehension and recall. *Reading Research Quarterly, 28*(3), 264-276.

Gee, J. (2007). *What video games have to teach us about learning and literacy*. New York, NY: Palgrave.

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.

Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research*, *19*(2), 144-161.

Mumin, Z. (2011). Cleaning up native English-speaking semantic and morphosyntactic confusion with Spanish *por* and *para*. *Journal of Linguistics and Language Teaching*, *2*(1), 195-214.

Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject-verb agreement processes in comprehension. *Journal of Memory and Language*, *36*(4), 569-587.

Peterson, M. (2010). Computerized games and simulations in computer-assisted language learning: A meta-analysis of research. *Simulation & Gaming, 41*(1), 72-93.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York, NY: Springer.

Qiao, X., Shen, L., & Forster, K. I. (2012). Relative clause processing in Mandarin: Evidence from the maze task. *Language and Cognitive Processes*, *27*(4), 611-630.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rayner, K. (1998). Eye movement in reading and information processing: 20 years of research. *Psychological Bulletin, 24*(3), 372-422.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning, 63*(3), 595-626.

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*(2), 223-247.

Rubio, F., & Thoms, J. J. (2012). Hybrid language teaching and learning: Looking forward. In F. Rubio, J. J. Thoms, & S. Katz Bourns, *Hybrid language teaching and learning: Exploring theoretical, pedagogical and curricular issues* (pp. 1-9). Boston, MA: Heinle Cengage Learning.

Scida, E. E., & Saury, R. E. (2006). Hybrid courses and their impact on student and classroom performance: A case at the University of Virginia. *CALICO Journal, 23*(3), 517-531.

Smith, B. (2003). Computer-mediated negotiated interaction: An expanded model. *Modern Language Journal, 87*(1), 38-57.

Sykes, J., & Reinhardt, J. (2013). *Language at play: Digital games in second and foreign language teaching and learning*. Upper Saddle River, NJ: Pearson.

Theriault, D. (2009). *The use of pictures in L2 vocabulary learning: Measuring the processing of forms and meanings* (Unpublished master's thesis). Concordia University, Montreal, Canada.

Thorne, S. L., Black, R. W., Sykes, J. M. (2009). Second language use, socialization, and learning in Internet interest communities and online gaming. *The Modern Language Journal, 93*(s1), 802-821.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27*(2), 173-204.

Toyoda, E., & Harrison, R. (2002). Categorization of text chat communication between learners and native speakers of Japanese. *Language Learning and Technology, 6*(1), 82-99.

Tremblay, P., & Gardner, R. C. (1995). Expanding the motivation construct in language learning. *The Modern Language Journal, 79*(4), 505-518.

VanPatten, B. (1987). Classroom learners' acquisition of *ser* and *estar*. Accounting for developmental patterns. In B. VanPatten, T. R. Dvorak, & J. F. Lee (Eds.), *Foreign language learning: A research perspective* (pp. 61-75). Cambridge, MA: Newbury House.

VanPatten, B. (2004). *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Lawrence Erlbaum.

VanPatten, B., & Cadierno, T. (1993). Input processing and second language acquisition: A role of instruction. *The Modern Language Journal*, *77*(1), 45-57.

VanPatten, B., Collopy, E., Price, J., Borst, S., & Qualin, A. (2013). Explicit information, grammatical sensitivity, and the first-noun principle: A cross-linguistic study in processing instruction. *The Modern Language Journal, 97*(2), 506-527.

Witzel, J., & Forster, K. I. (2014). Lexical co-occurrence and ambiguity resolution. *Language, Cognition, and Neuroscience, 29*(2), 158-185.

Witzel, N., Witzel, J., & Forster, K. I. (2012). Comparisons of online reading paradigms: Eye tracking, moving window, and maze. *Journal of Psycholinguistic Research, 41*(2), 105-128.

Appendix

Questions on the Story Maze Task Reaction Questionnaire

Students rated their experience from *5* (*strong yes*) to *1* (*strong no*) for #1-14, and #15 was open-ended.

1. Did you find this task fun?
2. Did you find this task helpful?
3. Did you find this task more enjoyable than online workbook assignments? In other words, would the maze task be a good supplemental activity to those assignments?
4. Did you find this task helpful for your Spanish learning?
5. Did you like the pictures presented with the sentences?
6. Did the pictures help you complete the sentences?
7. Did you like how the sentences were related to each other in a story-like manner? (That is, do you think this helped keep your interest?)
8. Did you find yourself *wanting* to try the sentence again if you got it wrong rather than just passing through it?
9. Do you think that this task can help others learn Spanish?
10. Do you think that this task could be helpful for other languages?
11. Do you think this type of practice carries over to doing better on exams/papers, etc.?
12. Did you think that the first session was just as fun as the fourth? (In other words: no it got old fast, or yes you think you could really get into it for a whole semester?)
13. Do you think that if there was a tally of reaction times (that is, how fast you are going), this would increase the fun factor of getting the answer correct?
14. Do you think that this would be a good addition to the Spanish curriculum?
15. Please write down any other comments/thoughts.