Task Type in Reading Span Tests Matters:
An Eye Movement Study*

Oya Özemir

Abstract

Reading span tests (RSTs) are commonly used to measure working memory (WM) capacity in L2 research. RSTs require simultaneous processing (secondary task) and storage (primary task) of information. However, the type of the secondary task varies from one study to another. Since syntactic and semantic processing in the L2 might be different, this study examined how the eye movements of late L2 learners changed when RSTs in the L2 involving syntactic accuracy vs. semantic plausibility judgment as the secondary task were administered. The findings indicated that eye movements displayed different patterns in semantically and syntactically anomalous sentences. Additionally, eye movement patterns of the participants were similar to those of native speakers reported in previous L1 (English) research in terms of processing semantic anomaly, but not syntactic inaccuracy.

Keywords: Working memory, Reading span tests, Secondary task, Eye movements, L2 sentence processing

Introduction

Reading Span Tests (RSTs) first developed by Daneman and Carpenter (1980) are one of the complex span measures that are predominantly used in research investigating working memory and reading comprehension relationship. As opposed to the simple span measures, which involve only the storage function of WM, complex span measures such as RSTs involve a dual-task paradigm in that both storage (i.e., primary task) and processing (i.e., secondary task) functions are represented in the test design.

In the original version of the RST (Daneman & Carpenter, 1980), participants read 60 unrelated sentences in sets of two to six sentences (13-16 words in length) aloud at their own pace (processing) and were asked to recall the last word of each sentence (storage) at the end of each set in the order in which they appeared. In order to prevent participants from focusing on the sentence-final words without giving much attention to processing the sentences, in a second experiment, Daneman and Carpenter (1980) asked the participants to indicate the veracity of the sentences they read by responding true or false. In this version of the test, the sentences were drawn from general knowledge quiz books and covered different domains, including the sciences, literature, geography, history, and current affairs, and were selected to be of moderate difficulty (e.g., You can trace the languages English and German back to the same roots).

Dr. Oya Özemir, Instructor, Yıldız Technical University, School of Foreign Languages, Davutpaşa Campus, Istanbul. oyaozemir@hotmail.com

Different versions of this test were used in L1 research. For example, in Hannon and Daneman's (2001) version of the task, participants were required to make judgments about the sensibility of each sentence as some of the sentences were implausible (e.g., The umbrella grabbed its bat and stepped up to the plate). Tirre and Pena (1992) used a task in which participants made judgments about the truth of sentences that were general knowledge statements about the natural and behavioral sciences (e.g., Neurosis is best treated by neurosurgery).The version of RST used by Waters and Caplan (1996) employed a plausibility judgment task in which the unacceptable sentences violated the thematic role requirements of the verb (e.g., It was the pillow that clenched the man). In the RST developed by Turner and Engle (1989), correct sentences made both semantic and syntactic sense, yet for incorrect sentences they reversed the order of the last four to six words.

One issue concerning the use of RSTs is the test takers' tendency to "develop idiosyncratic strategies for balancing the processing and storage components" (Friedman & Miyake, 2004, p. 137). Studies in L1 where the participants were interviewed about their strategy use have demonstrated that these participants actually utilized strategies such as making associations to or between the to-be remembered words, rehearsing the to-be remembered words while reading the sentences or instead of paying attention to the meaning of the sentence concentrating on the to-be remembered word (Friedman & Miyake, 2004; Kaakinen & Hyönä, 2007). Strategy use has also been investigated in L1 research by tracking eye movements of the participants. Eye movement methodology, which involves displaying sentences on a computer screen and recording the readers' eye movements and fixations while they are reading, provides data for cognitive processes as eye movements are moment-to-moment indicators of the ease or the difficulty which readers experience Moreover, it is a nonintrusive method in that it does not interfere with the natural test taking or reading process and thus it has been preferred in strategy use studies.

Literature Review

Issues in Investigating the Role of WM in L2 Reading

Like L1 research, RSTs used in L2 studies to measure WM capacity showed variability in the type of secondary task. While some studies utilized grammaticality judgment (Alptekin & Erçetin, 2010; Harrington & Sawyer, 1992), others involved semantic plausibility judgment (Leeser, 2007; Walter, 2004). When the L2 population is considered, the variability in the nature of the secondary task (syntactic accuracy or semantic plausibility judgment) gains special importance because processing in the L2 has been proven to be qualitatively different than that in the L1 (see Clahsen & Felser, 2006; Roberts, 2010 for reviews). These differences have been highlighted both in morphosyntactic and semantic processing from a neurobiologically-based viewpoint in Ullman's Declarative/Procedural (DP) Model (Ullman, 2001a, 2001b, 2004, 2005).

According to the DP model, idiosyncratic (lexical) knowledge is learned and processed in declarative memory whereas learning and processing of grammatical rules generally depend on procedural memory in the L1. However, given that declarative memory improves during childhood, plateaus in adolescence and early adulthood and that procedural memory may attenuate during childhood, late L2 learners are expected to rely more on declarative memory for the grammatical functions which are learned and processed through the procedural system in the L1. In other words, as opposed to the L1, declarative memory subserves the learning and use not only of idiosyncratic lexical knowledge but also of complex linguistic forms and rules in late L2 acquisition. This may take the form of "lexicalizing" input of an essentially syntactic nature through chunking processes, semantic processing that relegates syntactic processing to secondary status, application of explicit rules, and associative generalizations. Thus, processing morphosyntactic units may take place as if they were individual lexical items instead of performing rule-governed combinatorial computations on them. To illustrate, instead of combining the verb "look" and the bound morpheme "-ed" sequentially while processing the past participle "looked", late L2 learners may process it as if it were a nondecompositional lexical item due to their having stored it as a chunk.

Nevertheless, it should be noted that according to the model, with enough exposure to language and with practice, some proceduralization of the grammar may be expected. (Ullman, 2001a, 2001b, 2005; Morgan-Short & Ullman, 2012).

Eye Movement Research in Anomaly Studies in L1

In order to find out where and when they produce disruption effects on sentence processing, both semantic/pragmatic and/or syntactic anomalies have been examined using eye tracking methodology (Braze, Shankweiler, Ni, & Palumbo, 2002; Ni, Fodor, Crain, & Shankweiler, 1998; Murray & Rowan, 1988; Rayner, Warren, Juhasz, & Liversedge, 2004).

Murray and Rowan (1998) investigated effects of pragmatic plausibility in sentences where they manipulated noun phrase and verb combinations in subject or object positions (1-4):

(1)   The hunters stacked the bricks. (plausible object)
(2)   The hunters stacked the tulips. (implausible object)
(3)   The bishops stacked the bricks. (implausible subject)
(4)   The bishops stacked the tulips. (implausible subject and object)

They divided the sentences into three regions for analysis: the initial NP, the verb, and NP2. They found immediate effects of subject-verb plausibility effect as there was first-pass reading effect on the subject NP and fixation duration on the following verb was significantly longer when the preceding NP was an implausible subject.

Tracking the eye movements of university students who were native speakers of English, Ni et al. (1998) compared sentences that contained no anomaly (5),

pragmatic anomaly (6), and syntactic anomaly (7). Reading time measures yielded no significant differences in the critical region (verb and the following word) across conditions. However, frequency of regressions produced significant differences at the critical region and the two subsequent regions. The syntactic anomaly condition induced more regressions than the pragmatic anomaly condition at the point of anomaly, which in turn induced more regressions than the control condition, across the three regions. On the other hand, the amount of regressions increased at the end of the sentence.

(5)   It seems that/ the cats/ won't usually/ eat the/ food we/ put on the porch.
(6)   It seems that/ the cats/ won't usually/ bake the/ food we/ put on the porch.
(7)   It seems that/ the cats/ won't usually/ eating the/ food we/ put on the porch.

The result that regressions, which could be interpreted as reanalysis of the problematic part of the sentence, were initially fewer in the pragmatic anomaly condition was attributed by the researchers to the likelihood that the readers continued to read to make sense of the sentence before attempting to reread it. While syntactic processing might be restricted with grammar rules, semantic processing might be more open-ended and readers might think an anomaly in meaning can be, for instance, a figure of speech, and thus is not an actual error.

Braze et al. (2002) investigated the eye movements of university students who were native speakers of English during the processing of sentences that contain syntactic anomaly and pragmatic anomaly. Their sentences were constructed in a similar fashion to the sentences in Ni et al. (1998) study and were divided into six regions. Region 1 consisted of the subject noun phrase and was one or two words long. Region 2 was a modal verb followed by an adverb. Region 3 consisted of the main verb and the word following. Regions 4 and 5 contained two words each. The sentence final region, region 6, contained one to three words. In the control condition (8), both the verbal inflection and meaning were congruous with preceding words. Both the syntactic and pragmatic anomalies were created in the main verb. In the pragmatic anomaly condition (9), the verb's semantic content did not match with the subject noun phrase. In the syntactic anomaly condition (10), the inflection of the verb conflicted with a preceding modal.

(8)   The wall / will surely / crack after / a few / years in / this harsh climate
(9)   The wall / will surely / bite after / a few / years in / this harsh climate
(10)   The wall / will surely / cracking after / a few / years in / this harsh climate

Having analyzed the reading times and regression patterns, Braze et al. (2002) concluded that "the sentence processor responds differently to each kind of anomaly" (p.38). Although both syntactic and pragmatic anomaly caused perturbations in eye movements, the patterns in the subsequent regions to the anomaly diverged. First-pass reading times for syntactic anomalies were marginally longer than non-anomalous sentences at the anomaly region. However, syntactic anomalies triggered many regressions initially, followed by a rapid return to baseline. Pragmatic anomalies resulted in lengthened reading times at the anomaly region, and this was followed by an

increase in regressions at the end of the sentence. Similar to Ni et al.'s (1998) findings, pragmatically anomalous sentences generated more regressions than control sentences or syntactically anomalous sentences beyond the point of anomaly. Therefore, Braze et al. (2002) confirm that there is a delay in the integration of pragmatic information.

Rayner et al. (2004) used sentences that described events in which an individual performed an action with an implement. The sentences had identical syntactic structure but differed in terms of whether a noun phrase contained in the sentences was an appropriate or inappropriate thematic-role filler for an accompanying verb. They selected noun phrase arguments for the verbs in three experimental conditions in such a way that when thematic assignment occurred at the critical target word, the sentence was either plausible (likely theme) (11), or implausible (unlikely theme given the implement used in the event) (12), or anomalous (inappropriate theme) (13).

(11)   John used a knife to chop the large carrots for dinner.
(12)   John used an axe to chop the large carrots for dinner.
(13)   John used a pump to inflate the large carrots for dinner.

The results indicated that there was immediate disruption caused by the target word (carrots) in the anomalous condition as illustrated by longer gaze duration on the word in that condition. However, disruption to reading occurred in late measures (i.e., the go-past measure) on the target word and in the gaze duration on the following word in the implausible condition. Anomalous violations also triggered more regressions in the posttarget region. Rayner et al. (2004) concluded that when a word was anomalous, it had an immediate effect on eye movements, but that the effect of implausibility was not as immediate.

In summary, although L1 research on the processing of syntactic and semantic anomaly has revealed mixed results, it has been demonstrated that syntactic or semantic anomaly affects fixation durations and/or regressive eye movements.

Eye Movement Research in L2

When compared to the research done in the L1, research in the L2 is extremely limited and does not concentrate on syntactic and semantic anomalies. Instead, in the L2, eye movement methodology has been used in bilingual syntactic processing studies, especially to understand what factors affect immediate syntactic processing in the L2 and to examine how processing evolves with increasing skill in the L2.

In studies investigating syntactic ambiguity resolution among late bilinguals, it was found that proficient bilinguals show immediate sensitivity to structural ambiguities created by prepositional attachment in the same manner as native speakers. However, their eye movement analyses showed that bilinguals had globally longer rereading times (Frenck-Mestre & Pynte, 1997). Moreover, it was found that when the bilingual's two languages differed in the processing of a given syntactic structure, changes were observed in parsing strategies over time as the bilingual gained experience with the L2 (Frenck-Mestre, 2002). Another result from the aforementioned research was that

transfer from the L1 (French) decreased as the time of exposure to the L2 (English) increased as bilinguals' processing became more native-like.

Insights from Eye Movements to Strategy Use in RSTs

Prompted by the concern that the test takers might develop a strategy of processing the sentences shallowly and focus on the to-be-remembered word, a few eye movement studies were conducted in the L1 to understand strategies employed by people who take the RST. Carpenter and Just (1989) registered eye movements of six low-span and six high-span participants during the administration of the original version of the RST. They found that participants who scored high on the span task read faster but spent more time on the sentence-final words than participants who scored low on the task. They explained that better readers read faster as they accessed the meaning of words faster and thus used their extra time for task-specific requirements. When they examined the gaze durations on the words in terms of length (indicator of word encoding) and frequency (indicator of lexical access), they found that low-span participants' gaze durations on low- and high-frequency words did not differ under no load (recalling two words) and load conditions (recalling five words). However, for high-span participants gaze duration was longer at low-frequency words than at high-frequency words under no load whereas it was the same under load condition. Carpenter and Just (1989) concluded that high-span participants used the strategy of processing sentences superficially to allocate their resources to remember sentence-final words as memory load increased. On the other hand, low-span participants continued reading sentences for meaning instead of adapting their cognitive system to increased load condition.

In order to investigate strategy use in traditional RSTs, Kaakinen and Hyönä (2007) used eye movement methodology to obtain a detailed view of the time-course of processing of the test materials. They used the experimenter-paced Finnish version of Daneman and Carpenter's (1980) RST in order to minimize the time participants could use to devise and implement memory encoding strategies. They analyzed gaze duration and total fixation times for the first word of the sentence, the words in the middle of the sentence, and the sentence-final "to-be-remembered" word in different memory load conditions (four and five set-size conditions). This way, they compared low and medium span participants whose maximum performance was at set-size four and high span participants whose maximum performance was at set-size five. Unlike the previous study (Carpenter & Just, 1989), they did not find any differences stemming from different on-line processing strategies employed by different span groups. Both the high-span group performing with set-size five and the other span groups in set-size 4 allocated less processing time to the irrelevant parts of the test sentences (beginning and middle region) and more time to the "to-be-remembered" word as the memory load increased. The aforementioned studies demonstrate that RST takers may not be behaving as the test designer intended them to. Although in RSTs participants are expected to process the sentences, they might well be ignoring this important component of the task and be giving attention to the storage part only. This is especially

true when there is no additional task that will ensure the processing (e.g., grammaticality or plausibility judgment).

Although there is eye movement research showing strategy use in RSTs in the L1 (Carpenter & Just, 1989; Kaakinen & Hyönä, 2007), we do not have information about the reading behavior of the participants when RSTs themselves require grammaticality or plausibility judgments for task resolution, in addition to storage in the L2. Therefore, whether an individual's RST performance would change or remain the same based on the type of processing task is intriguing, particularly in the case of L2 learners. The present study was undertaken to investigate this issue.

Research Questions and Hypotheses

This study examines the eye movement patterns of late L2 learners[1] as they completed RSTs in the L2 in order to determine whether the type of linguistic secondary task employed affects their processing of sentences. Since eye tracking methodology provides moment-to-moment information of on-line processing decisions, it was thought to be appropriate as a validation method. The following research question was investigated:

Do the eye movements show differences when the RST administered in the L2 involves syntactic accuracy vs. semantic plausibility as the secondary task?

The secondary task of the RSTs were designed to ensure processing by creating disruptions in the normal reading and thus test takers were expected to show processing difficulties in the disruptive regions in their eye movements. Although research in the L1 has shown that syntactic or semantic anomaly affected fixation durations and/or regressive eye movements (Clifton, Staub, & Rayner, 2006), the results are mixed. What generally defines the eye movement patterns of native speakers encountering a semantic or syntactic problem is either longer first-pass reading times (Braze et al., 2002; Murray & Rowan, 1998; Rayner et al., 2004) or regressions (Braze et al., 2002; Ni et al., 1998). Consequently, in this study, critical regions with semantic or syntactic anomaly are expected to cause longer fixations and more regressions compared to control sentences (Hypothesis 1).

The second hypothesis was formed based on the DP Model. In other words, since the participants of the current study are late L2 learners who began to learn English in the middle childhood or later in formal settings, they may be relying primarily on their declarative memory instead of procedural memory in processing L2 syntax. Given the predictions of the DP Model, it was hypothesized that semantic plausibility judgments would yield eye movements similar to those of native speakers reported in the literature (e.g., Braze et al., 2002) while syntactic accuracy judgments

---

[1] In line with van Hell and Tokowicz (2010), late L2 learners in this study are defined as individuals who "learned their L2 in middle childhood (around 8-10 years) or later, well after adequate L1 language skills had been achieved" (p. 44).

would not (Hypothesis 2). Specifically, L1 research that this study was modeled after has shown that semantic anomaly and syntactic anomaly cause longer first-pass reading times in the critical region of anomalous sentences than in control sentences (Braze et al., 2002)[2] . However, while regressions originate from the last region in sentences containing semantic anomaly, they normally originate from the region where the verb is located in sentences containing syntactic anomaly (Braze et al., 2002).

Participants

Thirty three undergraduate students (28 female, 5 male) majoring in English Language Teaching (ELT) in the Foreign Language Education Department at a state university took part in the study. All participants had normal or corrected-to-normal vision. Participants' ages ranged from 20 to 22, with a mean of 20.6. Their years of exposure to English ranged from 9 to 13 years, with a mean of 10.5 years. Therefore, it can be said that they started learning English in middle childhood. As individuals who become exposed to L2 "in middle childhood (around 8−10 years) or later", these participants can be categorized as "late L2 learners" (van Hell & Tokowicz, 2010, p.44). They had been successful on the university's English proficiency test, whose minimum pass mark is accepted as the equivalent of 79 on the TOEFL IBT. They could be considered moderately proficient in English. They received course credit for their participation in the study.

Materials

Two RSTs in the L2 (English) was developed; these tests differed in the type of secondary task used (syntactic accuracy vs. semantic plausibility). Each RST consisted of 42 unrelated sentences and had four sets, each of which contained three trials (Alptekin & Erçetin, 2010; Harrington & Sawyer, 1992; Kaakinen & Hyöna, 2007). The sentences were generally constructed in line with the design in Braze et al. (2002) study. As delineated in Chapter 3, the sentences in Braze et al. (2002) were divided into six regions for the eye movement analysis. Yet one problem with Braze et al.'s (2002) study is the unequal length of the regions and of the words in the target region. Although they applied length correction procedures, these procedures are thought to be less reliable when the target region is short (Rayner et al., 2004), as is the case with Braze et al.'s (2002) study. Therefore, in the present study regions were divided in such a way that each region contained 11 to 13 characters and each region contained meaningful syntactic units, and thus they were reduced to four. Like Ni et al.'s (1998) and Braze et al.'s (2002) experiments, the syntactic inaccuracy and semantic implausibility were created in the main verb and this target word was equal in length (five characters). Words in the sentences were roughly of the same frequency as well[3] .

---

[2] It should be noted that other similar studies (Murray & Rowan, 1998; Ni et al., 1998) did not find significantly longer first-pass reading times in the critical region unlike Braze et al. (2002).
[3] www.wordcount.org which presents the 86,800 most frequently used English words taken from British National Corpus was used to check the frequencies. Words used in the tests were chosen from the first 10.000 most frequent words.

In control sentences (14), Region 1 consists of the subject noun phrase. Region 2 is a modal auxiliary verb followed by an adverb. Region 3 consists of the transitive main verb and its direct object. Region 4 contains a prepositional phrase or an adverbial clause of purpose. Both the verbal inflection and meaning are congruous with the preceding auxiliary in control sentences. In the syntactic inaccuracy condition (15), the inflection of the verb conflicts with a preceding auxiliary. In the semantic implausibility condition (16), the verb's semantic content does not match the following noun phrase used as a direct object. The following are some sample sentences used:

(14)   The residents / should never / allow the dogs / around the pool
(15)   The little kid / has severely / broke his leg / at a nearby park
(16)   The entire team / will equally / taste the costs / for their camp

Apparatus

Eye movements were recorded by the Applied Science Laboratories' (ASL) D6 Desk Mounted Optics remote eye tracker. The system consists of a control unit, a subject display monitor, a D6 Optics module and an interface PC. The D6 Optics module is placed on a stand directly under a nineteen-inch subject display monitor which displays the stimulus. The eye-tracking camera connects directly to the interface PC and is positioned to the left of the optics module. There is also a head-tracking camera positioned to the right of this module. This camera moves automatically with the readers' head movements and thus it is not necessary to use a chin rest or bite bar which will be obtrusive for the natural reading process.

Procedure

After having completed a demographic questionnaire, the participants were told about the procedure and how the camera worked. The participant sat in front of the subject display monitor at an optimal distance which is approximately 21 inch (61 cm) from the camera. Each session began with calibration.

The tests were administered randomly on two separate days. When the experiment started, the participants first read the instructions explaining the steps of the test and were encouraged to ask questions if they did not understand the instructions. The test started with a practice session and the data were recorded from the practice session to the end of the test. The participants were tested on the same set of sentences. Each sentence was displayed on the screen for seven seconds. This time restriction was deemed necessary to prevent application of rehearsal strategies which might influence the RST scores. The participants read the sentences in sets of three trials and pressed a button indicating whether the sentence was correct or not while reading. After having finished each set, the participants were asked to recall the final words of each sentence in the set and to type it in a box appearing on the screen.

Data Analysis

The analyses were conducted using the ASL Results program. Fixation sequence and duration are available on the program. Fixations less than 80 ms were automatically eliminated and fixations greater than 1000 ms were not counted during the analyses (Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007; Traxler, Pickering, & Clifton, 1998).

In addition, sentences with track losses were excluded and data from some participants were not usable due to extreme losses in the corneal and pupil reflections during the experiment. Consequently, four participants' data from the syntactic inaccuracy test, seven participants' data from the semantic implausibility test were excluded.

Measures

Three standard measures of eye movements were analyzed:
1. First-pass reading time: This is defined as the summed duration of fixations in a region from first entering it to first leaving it, either to the left or to the right.
2. Second-pass reading time: This is defined as the summed duration of all fixations made in a region after going past the region and then regressing back into it.
3. First-pass regression outs: These are defined as the number of eye movements in which a first pass fixation in a region was followed by a regression to an earlier region.

These measures were calculated for each of the four regions manually. First-pass and second-pass time durations in four regions were analyzed by ANOVAs. First-pass regression outs were also analyzed using ANOVA.

Results

First-Pass Reading Time

Table 1 provides the descriptive statistics for the first-pass reading time in seconds for control and anomaly sentences across the four regions.

As can be seen from Table 1, the mean of reading times across sentence types within each region does not vary much except Region 4 where reading times for control and semantic implausibility sentences are considerably higher than those of syntactic inaccuracy.

A 3 (sentence type) x 4 (region) repeated measures ANOVA was conducted to examine whether first-pass reading times significantly differ across different sentence types across regions. Mauchly's Test of Sphericity was significant ($p < .01$) for the sentence and region main effects as well as the sentence x region interaction effect. Therefore, Greenhouse-Geisser correction factor (Epsilon=.68, .71, .61 respectively) was used to evaluate the observed F ratios. Significant main effects for sentence type,

$F(1.34, 32.38) = 4.41$, $p > .05$, $\eta^2 = .15$ and region $F(2.13, 51.19) = 7.21$ $p > .01$. $\eta^2 = .23$ were attained. There was also a significant sentence type by region interaction effect $F(3.64, 87.36) = 16.21$, $p > .01$, $\eta^2 = .40$.

Table 1. Descriptive statistics for first-pass reading time (sec)

Sentences	M	SD	N
Region 1			
Control	8.362	2.802	25
Semantic implausibility	8.707	3.778	25
Syntactic inaccuracy	8.047	3.641	25
Region 2			
Control	7.581	2.141	25
Semantic implausibility	7.239	2.461	25
Syntactic inaccuracy	7.416	2.673	25
Region 3			
Control	8.176	2.567	25
Semantic implausibility	7.554	2.748	25
Syntactic inaccuracy	8.313	3.345	25
Region 4			
Control	10.442	3.062	25
Semantic implausibility	11.081	3.951	25
Syntactic inaccuracy	6.793	3.075	25

In order to further probe the interaction effect, a one-way repeated measures ANOVA was conducted for each region to understand whether the three types of sentences show significant differences. Region was held constant.

The ANOVA did not yield a significant difference across sentences in Region 1, $F(2, 48) = 0.86$, $p > .05$, in Region 2, $F(2, 48) = 0.33$, $p > .05$, and in Region 3, $F(2, 48) = 1.42$, $p > .05$. On the other hand, a significant effect of sentence type was observed in Region 4, $F(1.47, 35.15) = 22.63$, $p < .05$ (with Greenhouse-Geisser correction, Epsilon = .73), $\eta^2 = .48$.

In order to further follow up the significant effect of sentence type in Region 4, Post hoc comparisons using the Bonferroni procedure were conducted. The results showed that control ($M = 10.44$, $SD = 0.61$) and semantically implausible ($M = 11.08$, $SD = 0.79$) sentences had significantly longer reading times than syntactically inaccurate ($M = 6.79$, $SD = 0.61$) sentences in this region.

Second-Pass Reading Time

Table 2 provides the descriptive statistics for the second- pass reading time in seconds for control and anomaly sentences across the four regions.

There is a decrease in second-pass reading times compared to first-pass reading times. The reading times for different types of sentence seem to be similar within each region. On the other hand, the reading times in Region 3 are longer compared to those in the other regions.

A 3 (sentence type) x 4 (region) repeated-measures analysis of variance (ANOVA) was conducted to examine whether fixation durations change for different sentence types in different regions in the second pass. Mauchly's Test of Sphericity was significant ($p < .01$) for the main effect of region and thus Greenhouse-Geisser correction factor (Epsilon=.65) was used to evaluate the observed F ratios. The assumption of sphericity was sustained ($p > .05$) for the main effect of sentence type and the sentence x region interaction. There was a significant main effect for region $F(3,46.65) = 35.61$, $p > .01$, $\eta^2 = .59$, but not for sentence type. There was also significant sentence by region interaction effect $F(6,144) = 6.24$, $p > .01$, $\eta^2 = .20$.

In order to follow up the interaction effect, a one-way repeated measures ANOVA was conducted for each region to determine whether the three types of sentences show significant differences. Region was held constant.

Table 2. Descriptive statistics for second-pass reading time (sec)

Sentences	M	SD	N
Region 1			
Control	1.460	1.132	25
Semantic implausibility	1.063	.787	25
Syntactic inaccuracy	.967	.997	25
Region 2			
Control	3.401	2.004	25
Semantic implausibility	2.266	1.467	25
Syntactic inaccuracy	3.701	2.072	25
Region 3			
Control	4.260	2.233	25
Semantic implausibility	3.288	2.192	25
Syntactic inaccuracy	4.030	2.398	25
Region 4			
Control	2.027	1.154	25
Semantic implausibility	2.293	1.901	25
Syntactic inaccuracy	1.538	1.161	25

The ANOVA yielded a nonsignificant effect of sentence type in Region 1, $F(2, 48) = 2.56$, $p > .05$ and in Region 4, $F(2, 48) = 2.17$, $p > .05$. However, a significant effect was observed in Region 2, $F(2, 48) = 6.88$, $p < .05$, $\eta^2 = .22$ and in Region 3, $F(2, 48) = 3.15$, $p < .05$, $\eta^2 = .10$.

In order to further follow up the significant effect of sentence type in Region 2 and Region 3, Post hoc comparisons using the Bonferroni procedure were conducted. In Region 2, both control ($M = 3.40$, $SD = 0.40$) and syntactically inaccurate ($M = 3.70$, $SD = 0.41$) sentences had longer reading times than semantically implausible ($M = 2.26$, $SD = 0.29$) sentences while there was no significant difference between the control and the syntactically inaccurate sentences. Pairwise comparisons for Region 3 showed that the only significant difference was between control ($M = 4.26$, $SD = 0.43$) and semantically implausible ($M = 3.28$, $SD = 0.43$) sentences. There were no significant differences between syntactically inaccurate ($M = 4.03$, $SD = 0.48$) sentences and control and semantically implausible sentences. In short, only control sentences had significantly longer fixation duration than the other sentences in this region.

First-Pass Regression Outs

Regressions are other indicators of processing difficulty that were used in this study. The number of regressions for each sentence was calculated for each participant. Since it was not possible to regress from the first region, only regressions from regions 2, 3, and 4 were calculated. The regression means are based on the total number of regressions per condition (control, syntactic inaccuracy, semantic implausibility) per subject, divided by the number of sentences in each condition.

Table 3 provides the average number of first-pass regression outs from control and anomaly sentences across regions. Regressions for sentence types seem to be similar within each region except for Region 4 where semantically implausible sentences have higher frequency of regressions than those of the other sentence types. This region also contains more regression-outs compared to the other regions.

A 3 (sentence type) x 3 (region) repeated-measures analysis of variance (ANOVA) was conducted on the number of regressions for control and anomaly sentences. The assumption of sphericity was met ($p > .05$) for the main effects of sentence type and region. Mauchly's Test of Sphericity was significant ($p < .00$) for the sentence x region interaction effect. Therefore, Greenhouse-Geisser correction factor (Epsilon=.51) was used to evaluate the observed F ratios. Both the main effect for sentence $F(2,42) = 26.53$, $p > .05$, $\eta^2 = .55$ and region $F(2,42) = 70.44$, $p > .01$, $\eta^2 = .77$ and the interaction effect $F(2.04,42.98) = 34.50$, $p > .01$, $\eta^2 = .62$ were significant. In order to follow up the interaction effect, a one-way repeated measures ANOVA was conducted for each region to determine whether the regressions show significant differences.

The ANOVA did not yield a significant effect of sentence type for Region 2, $F(2, 50) = 1.18$, $p > .05$. On the other hand, a significant effect was observed in Region 3, $F(1.36, 31.20) = 22.02$, $p < .05$ (with Greenhouse-Geisser correction, Epsilon = .68), $\eta^2 = .49$ and in Region 4, $F(1.55, 34.09) = 38.24$, $p < .05$ (with Greenhouse-Geisser correction, Epsilon = .73), $\eta^2 = .64$.

In order to further follow up the significant effect of sentence type in Region 3 and Region 4, Post hoc comparisons using the Bonferroni procedure were conducted. In Region 3, most of the regressions out occurred in the syntactically inaccurate sentences ($M = 4.29$, $SD = 2.85$). This was followed by the control sentences ($M = 2.29$, $SD = 1.50$) and semantically implausible sentences ($M = 1.08$, $SD = 2.85$).

Table 3. Descriptive statistics for the number of first-pass regression outs

Sentences	M	SD	N
Region 2			
Control	.681	.824	22
Semantic implausibility	.909	1.444	22
Syntactic inaccuracy	1.000	1.976	22
Region 3			
Control	2.409	1.608	22
Semantic implausibility	1.136	1.552	22
Syntactic inaccuracy	4.409	2.954	22
Region 4			
Control	4.272	2.649	22
Semantic implausibility	11.727	3.548	22
Syntactic inaccuracy	6.545	4.553	22

In Region 4, there were more regressions from the semantically implausible sentences ($M = 11.26$, $SD = 4.13$) than syntactically inaccurate sentences ($M = 6.30$, $SD = 4.60$) and control sentences ($M = 4.09$, $SD = 2.74$).

These findings are partly consistent with the research hypotheses. Although Hypothesis 1 maintained that critical regions in the anomalous sentences should yield disruption effects which could be demonstrated by longer reading times and more regressions than control sentences, this disruption effect was observed elsewhere in the first- and second-pass reading times. On the other hand, first-pass regressions indicated an immediate disruption effect in syntactically inaccurate sentences in the critical region (Region 3) and Region 4 contained significantly more regressions in semantically implausible sentences. These results are consistent with the previous research in the L1 (Braze et al., 2002; Ni et al., 1998).

Furthermore, Hypothesis 2 that predicted a similar pattern of eye movements to that of native speakers for the semantically implausible sentences but a different pattern for the syntactically inaccurate sentences was partly confirmed. According to the first-pass reading time results, neither semantic implausibility nor syntactic inaccuracy revealed longer first-pass reading times in the critical region unlike Braze et al.'s (2002) findings. On the other hand, first-pass regression out results were similar to those found in the L1 research. Like the previous research with L1 speakers (Braze et al., 2002), semantic implausibility caused regressions in Region 4. Again, like Braze et al.'s (2002) study, regressions in syntactically inaccurate sentences occurred in the critical region.

Discussion

The findings indicate that semantically and syntactically anomalous sentences yield different processing patterns for not only native speakers as demonstrated in the L1 research but also for nonnative speakers. This result emphasizes the importance of the secondary task in RSTs.

To begin with, for the first-pass reading time, the participants seemed to process the sentences in both RSTs in a similar fashion until the last region. Longer reading times attained in Region 4 in control and semantically implausible sentences are not unlikely because this is the region including the word to be recalled. In fact, previous eye tracking research in the L1 showed that participants spent longer time on the sentence-final word (Carpenter & Just, 1989; Kaakinen & Hyönä, 2007).

In fact, the result that Region 4 does not induce as long a reading time in syntactically inaccurate sentences as in control and semantically implausible sentences suggests that the participants did not adopt a strategy to concentrate on the final word for all sentence types. Actually, the longer reading times in control sentences may indicate that the participants, not having had difficulty in processing the control sentences, were likely to have allocated more time to storing the last word in the sentence. Although, the same can be said for semantically implausible sentences, previous research in the L1 has shown that reading times increase towards the end of sentences in such anomaly conditions (Ni et al., 1998). Another explanation for the longer reading times in Region 4 in semantically implausible sentences may be the sentence wrap-up effect, which claims that readers fixate longer at the end of sentences to integrate information and to clarify inconsistencies or ambiguities they could not resolve within the sentence (Just & Carpenter, 1980).

On the other hand, according to second-pass reading times the only significant difference was attained in Region 2 in control and syntactically inaccurate sentences. Although this is not the critical region, longer reading time in this region is not unexpected because the ungrammaticality in this type of sentence is based on the incongruity between Regions 2 (e.g., will probably) and 3 (e.g., going abroad) and it is likely that the participants in the study reread the second region to be sure of the anomaly.

Another significant difference between the sentence types in Region 2 was attained in control sentences, but the reading time in this condition is not as high as that in syntactic inaccuracy sentences. Having difficulty in an unautomatized syntactic processing, the participants might have felt the need to check Region 2 even in control sentences. Therefore, it can be said that syntactically inaccurate sentences induced the most rereadings among all sentence types. When first-pass regression patterns are examined, it can be seen that syntactically inaccurate sentences generated significantly more regressions out of Region 3. This immediate disruption effect shows that the participants experienced difficulty in the critical region and needed to reread Region 2 as demonstrated by longer reading in that region. The results of first-pass reading times

seem to support this view. That is, the lack of significantly longer reading times in Region 4 in syntactically inaccurate sentences may indicate that the participants could not spend much time for storage, probably because they had more difficulty in processing syntactic inaccuracy rather than the semantic implausibility. This kind of disruption and reanalysis suggest that participants were not able to discern the problem through procedural (implicit, nonconscious) ways, thereby examining sentence constituents in an associative way, the way declarative memory would subserve.

As for semantic processing, neither during the first-pass nor the second-pass a significantly longer reading time was observed in the critical region. Unlike in the syntactic inaccuracy condition, the critical region in the semantic implausibility condition is confined to Region 3. While syntactic parsing is sequential, semantic processing is based on thematic relations (rather than sequential relation). Therefore, a lack of disruption in the region prior to the critical region is not expected whereas the critical region is prone to manifest disruption effects.

When the results of this study are compared to the previous research on syntactic and semantic anomaly processing in the L1, both similarities and differences can be seen. First, the sentences in this study were structurally similar to those in Ni et al. (1998) and Braze et al. (2002) and thus these studies provide a direct comparison with the native speaker population. The lack of a significant difference between control and experiment sentences or between syntactically inaccurate and semantically implausible sentences in the critical region in the first-pass seems unexpected, but it was a finding in the study by Ni et al. (1998) as well. They could not find a difference between the two types of anomaly in the first-pass reading time in the anomaly region or the preceding region. The experiment sentences did not significantly differ from control sentences in the critical region, either. On the other hand, Braze et al. (2002) found longer first-pass reading time in the critical region in the semantic and to a lesser extent in the syntactic anomaly sentences. Nevertheless, it should be pointed out that semantic incongruity in these studies emerges from incongruity between the subject and the verb (e.g., The wall will surely bite after a few years in this harsh climate), not from that between the verb and object as was the case in this study.

Similar to Braze et al. (2002), Murray and Rowan (1998) found significant first-pass reading effects in sentences which included subject-verb implausibility (e.g., The bishops stacked the bricks), but not in sentences which included verb-object implausibility (e.g., The hunters stacked the tulips). In this study, too, semantic implausibility stems from the thematic relationship between the verb and the direct object following it. Therefore, it can be said that L2 learners in this study behaved like native speakers who did not display longer reading times in the critical region (Murray & Rowan, 1998; Ni et al., 1998) and who regressed more beyond the point of the critical region (Braze et al, 2002; Ni et al., 1998) in the case of semantically driven RSTs.

Second, problems in syntax triggered regressions initially at the critical region both in Ni et al. (1998) and Braze et al. (2002) while regressions started later in the

sentence in the semantically anomalous sentences. This result was also obtained in the present study. Except for these similarities, there is an additional result in that L2 learners in this study reread Region 2 in syntactically inaccurate sentences. Actually, a direct comparison with the native speakers here is not possible because neither Ni et al. (1998) nor Braze et al. (2002) reported second-pass reading time. Nevertheless, landing sites for regressions in Braze et al. (2002) did not significantly differ in the syntactically anomalous sentences. Thus, native speakers in their study cannot be said to have returned to a prior region to reread that section significantly longer than other regions.

Conclusion

To conclude, it can be said that the participants actually processed the sentences and responded to the syntactic or semantic problems in the sentences instead of just focusing on the sentence-final word as is a common tendency in RSTs. In addition, eye movements followed different patterns in sentences that required primarily semantic and primarily syntactic processing. The participants' processing showed a lack of proceduralization of syntax and was primarily guided by lexical/semantic processing choices even in sentences where syntactic judgments were required. Besides, while they processed the sentences in a similar way to the native speakers in the semantically driven test, their processing pattern diverged from that of the native speakers due to possibly the lack of proceduralized syntax in the syntactically driven test. As a result, it stands to reason that since L2 syntactic processing seems to be qualitatively different than L1 syntactic processing as shown by dissimilar eye movement patterns to those of native speakers in syntactic processing, using an RST that employs syntactic accuracy judgment in the L2 may not be appropriate for late L2 learners who have not proceduralized L2 syntax fully. Instead, an RST incorporating semantic plausibility judgment may involve less cognitive load for late L2 learners.

References

Alptekin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading, 33*(2), 206–219.

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research, 31,* 25–44.

Carpenter, P. A., & Just, M. A. (1989). The role of working memory in language comprehension. In D. K. Klahr and K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 31–68). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27,* 3–42.

Clifton, C., Jr., Staub, A., & Rayner, K. (2006) Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S.Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Oxford: Elsevier.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Frenck-Mestre, C. (2002). A on-line look at sentence processing in a second language. In R. Herrida & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 217–236). North Holland: Elsevier Science.

Frenck-Mestre, C. & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology, 50A,* 119–148.

Friedman, N. P. & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, *51,* 136–158.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93*(1), 103–128.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25–38.

van Hell, J. G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research, 26*, 43–74.

Just, M.A., & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329-354.

Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory, 15*(6), 634–646.

Leeser, M.J., (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: topic familiarity and working memory. *Language Learning, 57,* 229–270.

Morgan-Short, K., & Ullman, M. T. (2012). The neurocognition of second language. In A. Mackey & S. Gass (Eds.), *Handbook of second language acquisition* (pp. 282–299). New York, NY: Routledge.

Murray, W. S., & Rowan, M.(1998). Early, mandatory, pragmatic processing. *Journal of Psycholinguistic Research, 27*(1), 1-22.

Ni,W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: eye movement patterns. *Journal of Psycholinguistic Research*, *27*, 515–539.
Rayner, K.,Warren, T., Juhasz, B. J., & Liversedge, S. P.(2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*, 1290–1301.

Roberts, L. (2010). Parsing the L2 input, an overview: Investigating L2 learners' processing of syntactic ambiguities and dependencies in real-time comprehension. In G. D.Véronique (Ed.), *Language, interaction and acquisition* [Special issue], (pp. 189–205). Amsterdam: John Benjamins.

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun–noun compounds. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition,33*(6), 1162–1169.

Tirre, W. C., & Pena, C. M. (1992). Investigation of functional working memory in the reading span test. *Journal of Educational Psychology*, *84*(4), 462–472.

Traxler, M.J., Pickering, M.J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39,* 558–592.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, *28*, 127–154.

Ullman, M. T. (2001a). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research, 30*(1), 37–69.

Ullman, M. T. (2001b). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition, 4*(1), 105–122.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*(1–2), 231–270.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory and practice* (pp. 141–178). Washington, DC: Georgetown University Press.

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics, 25(3),* 315–339.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology, 49A*(1), 51–79.

Okuma Aralığı Testlerinde Görev Türü Önemlidir: Bir Göz Hareketi Çalışması

Özet

Okuma Aralığı Testleri (OAT) ikinci dilde (D2) yapılan araştırmalarda işler bellek kapasitesini ölçmede sıkça kullanılır. OAT'ler, işlem yapma (ikincil görev) ve bilgi depolama (birincil görev) işlerinin aynı anda yapılmasını gerektirir. Ancak, ikincil görev türü bir çalışmadan diğerine farklılık gösterir. D2'de sözdizimsel ve anlamsal aykırılıkların işlemesi farklılık gösterebileceği için, bu çalışmada D2'yi geç yaşta öğrenmeye başlayanların göz hareketlerinin OAT'ler D2'de ikincil görev olarak sözdizimi ya da anlam doğruluğunun yargılanmasını içerdiğinde nasıl değişeceği incelenmiştir. Sonuçlar, göz hareketlerinin sözdizimsel ve anlamsal aykırılık içeren cümlelerde değişiklik gösterdiğini işaret etmiştir. İlaveten, katılımcıların göz hareketleri yapılarının daha önce D1 olarak İngilizce'de yapılan araştırmalardaki katılımcıların göz hareketleri yapılarına anlamsal aykırılıkların işlemesinde benzediğini ama sözdizimi hataları işlemesinde benzemediğini göstermiştir.

Anahtar sözcükler: İşler bellek, Okuma aralığı testi, İkincil görev, Göz hareketleri, İkinci dilde cümle işleme.