

RUNNING HEAD: Island Constraints in Second Language Processing

Filler-gap Dependencies and Island Constraints in Second Language Sentence Processing

Akira Omaki

University of Maryland

Barbara Schulz

University of South Carolina

Address for correspondence:

Akira Omaki

Department of Linguistics

1401 Marie Mount Hall

University of Maryland

College Park, MD 20742

Phone: 301-405-8306

Email: omaki@umd.edu

Abstract

Second language (L2) processing may be different in various ways from processing in a native language, and it has been argued that one major difference is that L2 learners can only construct shallow representations that lack structural details (Clahsen & Felser, 2006). The present study challenges this hypothesis by comparing the extent to which advanced Spanish-English L2 learners and English native speakers make use of the relative clause island constraint in constructing filler-gap dependencies. In off-line acceptability judgment and on-line self-paced reading measures, the L2 group and native speaker control group demonstrate clear evidence for application of the relative clause island constraint. Our findings suggest that advanced L2 learners not only build abstract structural representations but also rapidly constrain the active search for a gap location, and cast doubt on the proposal that L2 learners are unable to build structural representations with grammatical precision.

FILLER-GAP DEPENDENCIES AND ISLAND CONSTRAINTS IN SECOND LANGUAGE
SENTENCE PROCESSING

Successful language comprehension by native speakers requires the parser to generate grammatically accurate structural representations of the incoming linguistic input. This paper is concerned with the nature of the representations generated by second language (L2) learners during real-time sentence processing, with a particular focus on whether L2 learners can build syntactic representations that support application of island constraints. The present study aims to contribute to the current debate on whether the L2 parser is sensitive to detailed grammatical information. Clahsen and Felser (2006) have proposed the Shallow Structure Hypothesis (henceforth SSH), which states that the L2 parser can only construct shallow structure representations that lack detailed syntactic information. Investigating the use of grammatical constraints like island constraints on long distance dependency formation provides a useful testing ground for Clahsen and Felser's hypothesis. Island constraints can be applied only when the abstract syntactic environment that defines islands is structurally represented, and therefore SSH predicts that L2 learners should violate island constraints in real-time processing. This article challenges SSH by presenting evidence that both native speakers and L2 learners respect island constraints and inhibit ungrammatical long distance dependency formation, suggesting that L2 learners can build abstract structural representations to guide the parser's active search for a gap position.

Filler-gap Dependencies in L2 Sentence Processing

A growing body of L2 sentence processing research has focused on the resolution of (temporary) syntactic ambiguity, and has tested whether L2 learners demonstrate the same parsing preferences as native speakers of a given target language. One type of ambiguity that has

recently received much attention in L2 sentence processing is filler-gap dependency processing in sentences that involve long distance dislocation of a constituent, as in *wh*-questions (1a) or relativization (1b).

- (1) a. *Which researcher* did John talk to ____ at the conference?
 b. This is *the researcher* that John talked to ____ at the conference.

In these constructions, the parser must identify the *gap* position (indicated by the underlines in (1)) in order to assign a thematic interpretation to the dislocated constituent (called the *filler*, indicated in italics in (1)). Previous studies that compared L2 learners' filler-gap processing to that of native speakers produced mixed results; L2 learners behaved in a native-like fashion in some respects but failed to show a native-like performance in others. For example, L2 learners seem to behave like native speakers in that they demonstrate evidence for *active gap filling*, i.e., the parser retrieves and structurally integrates the filler at the earliest potential gap position to assign a thematic interpretation, despite the possibility that the true gap location may be present in the upcoming part of the sentence. It has been hypothesized that the active search for a gap is driven either by the need to reduce the cost of retaining the filler in memory (Gibson, 1998) or by a processing principle that requires the parser to satisfy grammatical dependency as soon as possible (de Vincenzi, 1991; Frazier, 1987; Pritchett, 1992). There is ample evidence for active gap search in L1 parsing (e.g., Aoshima, Phillips, & Weinberg, 2004; Crain & Fodor, 1985; Frazier, 1987; Garnsey, Tanenhaus, & Chapman, 1989; Lee, 2004; Stowe, 1986; Sussman & Sedivy, 2003; see Phillips & Wagers, 2007 for a review of filler-gap dependency processing in native speakers), but evidence for active gap filling has been attested in L2 processing as well.

Williams, Möbius, and Kim (2001), for example, tested how English native speakers and Korean, Chinese and German learners of English process sentences like (2) by using a word-by-word reading task in which the participants had to make a “stop-making sense” decision if the sentence started to sound odd (Boland, Tanenhaus, Garnsey, & Carlson, 1995). In this sentence, the true gap location is after the preposition *into*, but the verb *push* is the earliest grammatical position where the *wh*-filler can be assigned a thematic interpretation. However, the filler was manipulated to be a plausible object of *push* in (2a) but an implausible object in (2b).

- (2) a. Which girl did the man push the bike into _____ late last night?
 b. Which river did the man push the bike into _____ late last night?

Williams et al. found that native speakers as well as L2 learners started making a stop-making sense decision at the verb in (2b), suggesting that the participants temporarily analyzed the *wh*-filler as the object of the verb and obtained an implausible interpretation (e.g., *pushed which river*). Moreover, reading times at the NP object region (*the bike*) were larger when the filler was a plausible object of the verb as in (2a), and this reading time slow-down suggests that the participants immediately analyzed the *wh*-filler as the object of the verb, and they had to reanalyze the sentence when the initial analysis was disconfirmed by the presence of the true direct object. These findings indicate that native speakers and non-native speakers actively predict the presence of a gap despite the lack of bottom-up evidence. Other L2 filler-gap processing studies have attested similar evidence for active gap processing behaviors in comprehension (Jackson & Dussias, 2008; Juffs, 2005; Juffs & Harrington, 1995; cf. Williams, 2006) as well as in production (Lieberman, Aoshima, & Phillips, 2006), suggesting that this is a

robust property of the L2 processing mechanism (see Dallas & Kaan, 2008 for a review of filler-gap dependency processing in L2 learners).

On the other hand, some studies on filler-gap processing have found evidence to support Clahsen and Felser's (2006) proposal that non-native speakers construct structurally shallower representations than native speakers. For example, Marinis, Roberts, Felser, and Clahsen (2005) examined processing of a cross-clausal *wh*-dependency (3) by English native speakers and advanced Chinese, Japanese, German and Greek learners of English, and found a contrast between the native speakers and non-native speakers in that the non-native speakers showed no clear reading time evidence for a pre-gap reactivation of the filler. They concluded from this finding that the L2 learners failed to postulate an intermediate trace (t') at the clause boundary.

- (3) The nurse who_1 the doctor argued t'_1 that the rude patient had angered t_1 is refusing to work late.

Felser and Roberts (2007) found a similar native vs. non-native contrast in a cross-modal picture priming study with native speakers of English and advanced Greek-English L2 learners. In reading sentences with ditransitive verbs like (4), the native speakers with a high memory span showed evidence for semantic reactivation of the filler when the probe was presented at the structurally defined gap position (gap probe), but not when the probe was presented in a position different from the gap position (pre-gap probe). On the other hand, the L2 learners showed semantic reactivation of the filler at both probe positions. Felser and Roberts interpreted the results to show that L2 learners resorted to keeping the filler constantly active in memory, rather than retrieving the filler at the structurally defined gap position.

- (4) Fred chased the squirrel to which the nice monkey explained the game's [*pre-gap probe*] difficult rules [*gap probe*] in the class last Wednesday.

Marinis et al. as well as Felser and Roberts interpreted these data to indicate that L2 learners do not postulate abstract representations like traces of the moved constituents, but rather form a direct lexical association between the filler and the lexical item (e.g., verb) that assigns thematic interpretations to the filler (Pickering & Barry, 1991).

These studies demonstrate behavioral differences between native speakers and L2 learners with respect to filler-gap dependency processing, but their findings are open to alternative interpretations. First, it is not clear that the experimental designs used by Marinis et al. (2005) and by Felser and Roberts (2007) were suitable methods for examining the nature of the representations constructed during real-time sentence processing, even for native speakers. For example, in the experimental design used in Marinis et al. and in the original study by Gibson and Warren (2004), the reading time differences found at the point of filler-retrieval can indicate whether the filler had been previously reactivated or not, but this previous reactivation does not necessarily implicate the presence of an abstract representation such as an intermediate trace, as the filler-retrieval and reactivation process could be achieved independently without postulating a trace (Pickering & Barry, 1991; for more discussion on this point, see Phillips & Wagers, 2007). As for Felser and Robert's study, McKoon, Ratcliff, and Ward (1994) demonstrated that the cross-modal priming effects reported in the L1 psycholinguistics literature are not consistently replicable, suggesting that it is difficult to rely on the cross-modal priming design as a measure of the representations that are constructed during real-time processing.

Second, findings by Marinis et al. and Felser and Roberts could indicate that the L2 learners are using non-target-like parsing procedures while building rich structural representations. For example, the lack of reading time evidence for intermediate reactivation in Marinis et al.'s study could indicate a ceiling effect in that the L2 learners were simply not efficient enough in processing the words in the critical regions or in retrieving the filler to demonstrate the expected reading time contrasts across conditions (DeKeyser, Schwartz, & Sprouse, 2006). The presence of associate priming at both probe positions in Felser and Roberts' study could indicate that L2 learners use their memory architecture in a different way than do native speakers. In fact, even the native speakers with low memory spans in Felser and Roberts' study did not show associate priming at the structural gap position, supporting the view that the properties of the memory system are a factor that complicates the distribution of priming effects.

Finally, it is possible that the L2 learners in Marinis et al. and Felser and Roberts had different grammatical knowledge than native speakers and nevertheless construct structural representations that are consistent with their own L2 grammars. Neither study provides an independent measure of whether the L2 learners have the relevant grammatical knowledge to demonstrate the expected processing behaviors. However, the hypothesized target-like reading time pattern in Marinis et al.'s study is predicated on the assumption that the participants have target-like grammatical knowledge of *wh*-movement rules and constraints, whereas Felser and Roberts' expected pattern of associate priming rests on the assumption that the participants have acquired the target-like structural representation of ditransitive constructions, which are known to show complex variations across languages (e.g., Pylkkänen, 2008). However, it is perfectly plausible that L2 learners do not have target-like knowledge in these domains. For example, studies using various off-line measures such as acceptability judgment tasks have suggested that

L2 learners do not always show native-like sensitivity to locality constraints on long distance movement (e.g., Hawkins & Hattori, 2006; Schachter, 1990; but cf. Belikova & White, 2009; Li, 1998; Martohardjono & Gair, 1993), and cross-linguistic differences in the structure of ditransitive verbs have led to many studies attesting non-target-like argument structure representations (e.g., Inagaki, 2001; Montrul, 2000; Whong-Barr & Schwartz, 2002). Given that the L2 learners in Marinis et al. and Felser and Roberts may not have the pre-requisite grammar for the target constructions, it is difficult to exclude the possibility that the native vs. non-native processing contrast is not due to differences in grammatical knowledge.

In summary, the previous arguments for SSH are amenable to alternative interpretations in that the non-target-like behaviors of the L2 parser may be unrelated to the nature of representations being constructed in real-time processing, or that the L2 learners simply may have had non-target-like structural representational options. Considering these alternative interpretations, it seems that an ideal test of SSH should have the following components: a) an off-line task with no time pressure, to examine whether L2 learners demonstrate the pre-requisite grammatical knowledge of the relevant rules and constraints, and b) an on-line experimental design that is able to elicit positive time course evidence that the rules and constraints are not deployed with grammatical precision in real-time language processing.

THE PRESENT STUDY

Our study further investigates structural details of filler-gap dependency processing while addressing the methodological concerns discussed above. Specifically, we use a) an acceptability judgment task to ascertain that our L2 learners have pre-requisite target-like grammatical knowledge in the relevant domain, and b) an experimental design that could elicit positive

reading time evidence for SSH. In order to achieve this, we examined off-line and on-line measures of the effects of island constraints on long distance dependency formation.

Long distance dependencies observed in English *wh*-questions or relative clauses are potentially unbounded in that the distance between the filler and the gap can be arbitrarily long, but it has been known since seminal work by Ross (1967) that there are syntactic domains called *islands* that are opaque to syntactic dependency formation. Subsequent studies have revealed cross-linguistic differences in some of the island domains (e.g., Rizzi, 1978; Stepanov, 2007), but the class of islands called *strong islands*, including, for example a relative clause island, is generally considered universal (cf. Engdahl, 1997; Kush, Omaki, & Hornstein, 2009), and various accounts have been proposed to provide a unified characterization of strong islands (e.g., Chomsky 1973, 1977, 1986; Huang, 1982; Lasnik & Saito, 1992; Rizzi, 1990; Uriagereka, 1999; for a review, see Szabolsci & den Dikken 2003). An example of relative clause (RC) island violation is illustrated in (5).

(5) * What did the reporter meet the politician [_{RC} who supported ____ at the congress].

In this example, the *wh*-filler *what* needs to be associated with the verb *supported* to receive a thematic interpretation, but since this verb is inside a relative clause island, the resulting dependency is ungrammatical. As we discussed above, the parser actively searches for a gap as soon as it encounters a filler. In (5) the sentence initial *what* is a filler and should consequently trigger an active gap search by the parser. Given that transitive verbs like *support* provide an object position that could serve as a gap site, one might expect that the parser should postulate an object gap as soon as it encounters this embedded verb. However, many sentence processing

studies have found that the native speakers immediately respect island constraints and do not attempt to postulate a gap inside an island (McElree & Griffith, 1998; Stowe, 1986; Traxler & Pickering, 1996; Yoshida, 2006). For example, an eye-tracking experiment by Traxler and Pickering (1996) used a plausibility mismatch design in (6).

- (6) a. We like the city / book that the author *wrote* unceasingly and with great dedication about _____ while waiting for a contract.
- b. We like the city / book that the author [_{RC} who *wrote* unceasingly and with great dedication] saw _____ while waiting for a contract.

Traxler and Pickering found a plausibility mismatch effect in (6a), i.e., the eye gaze duration at the verb *wrote* increased when the filler was an implausible object of the verb (i.e., *the city*), which provides another form of evidence that the parser actively associates a filler with the earliest potential gap position (see also Garnsey, Tanenhaus, & Chapman, 1989). However, the plausibility mismatch effect disappeared when the critical verb *wrote* was embedded inside an RC island (6b), despite the fact that the same verb was still linearly the first potential gap host after the filler. The absence of active gap creation inside an island suggests that the parser applies the island constraint in real-time processing and inhibits ungrammatical long distance dependency formation.

Let us now consider the prediction of SSH for this paradigm. For this purpose, it is useful to illustrate the exact representations that Clahsen and Felser propose for the sentences tested in Marinis et al.'s study (p.32, Clahsen & Felser, 2006). The representation that Clahsen and Felser

propose for native speakers is shown in (7a) with a slight modification, and the representation that is attributed to L2 learners is shown in (7b).

- (7) a. [_{DP} The nurse [_{CP} [*who*_i] the doctor argued [_{CP} [*t'*_i] that the rude patient had angered [*t*_i]]]] is refusing to work late.
- b. [The nurse] who [the doctor] argued [that [the rude patient] had angered] is refusing to work late.

What is important for the present study is that the structural representation attributed to L2 learners (7b) lacks a representational unit for a relative clause, *the nurse who...* Thus, if L2 learners only construct shallow representations as proposed by Clahsen and Felser, it is predicted that L2 learners should not respect the RC island constraint because there is no relative clause representation in their analysis. The paradigm in (6) presents an ideal testing ground for SSH in that unlike the studies by Marinis et al. or Felser and Roberts, the current design could potentially elicit positive, stronger time course evidence for SSH. If L2 learners actively search for gaps while only constructing shallow representations without syntactic details, then the RC island domain cannot be properly represented, and active gap filling (and consequently the plausibility mismatch effect) should be observed in non-island contexts (6a) and island contexts (6b) alike. On the other hand, if L2 learners turn out to be capable of building relative clause representations that define an RC island, then it is predicted that the plausibility mismatch effect should be observed in (6a) but not in (6b), replicating the native speaker results from Traxler and Pickering (1996). Note also that the application of island constraints is orthogonal to the issue of how L2 learners represent gaps (Felser & Roberts, 2007; Marinis et al., 2005). What the island

constraints do is to restrict the domain in which the parser searches for a gap, and hence the choice of representation alternatives (i.e., traces or direct lexical association) should not affect the expected reading time pattern in this design.

Taken together, investigations of active search for a gap and its interaction with island constraints can shed light on the nature of the linguistic representations that are constructed during real-time language processing. The current study uses the plausibility mismatch paradigm in (6) to test whether advanced Spanish-English L2 learners can construct a structural representation for an RC island and consequently constrain their active gap search process.

In addition, given the possibility that L2 learners could have non-target like grammatical knowledge with respect to island constraints, the present study examines whether the L2 learners in our study have the pre-requisite grammar using an off-line acceptability judgment task. On-line measures such as reaction times are notoriously susceptible to noise and are also likely to reflect a complex interaction of multiple cognitive processes (e.g., word recognition, lexical retrieval, motor planning, etc.), but in off-line acceptability judgment tasks the effect of time pressure and processing limitations can be alleviated to some extent (Chomsky, 1965; Schütze, 1996). Of course, showing that L2 learners have the pre-requisite grammar in off-line tasks is necessary but not sufficient for rejecting SSH, since it is possible that the detailed grammatical knowledge is used only in a second-pass parsing stage but not in first-pass processing (Townsend & Bever, 2001). Thus, combining off-line and on-line measures is crucial in testing whether L2 learners have island constraints and are able to deploy them properly in real-time processing.

Method

Participants

We recruited 56 participants from the University of South Carolina and University of Maryland communities, specifically, 32 native speakers of English and 24 advanced Spanish-speaking learners of English from South America or Spain. They all received course credit or were paid \$10 for their participation.

We examined the L2 learners' overall English proficiency by administering a c-test, which tests general language proficiency based on multiple deletions of parts of words in continuous texts (for details, see Eckes & Grotjahn, 2006). The L2 learners' average score was 42.5 out of 60 points ($SD = 10.1$), ranging from 24 to 58. We did not administer the c-test with the native speakers in the current study, but according to Schulz (2006) who reports data from 30 native speakers of English who took the same c-test, the average score was 50/60 ($SD = 7.7$), ranging from 26 to 59. Given that the score range of a representative sample of native speakers and that of our L2 learners is roughly equivalent, we concluded that our L2 group consists of highly proficient learners of English. The L2 learners' mean age of first exposure to English was 11.0 years ($SD = 3.4$) and they had received an average of 10.4 years of instruction ($SD = 5.0$).

Materials

Self-paced reading task stimuli. The materials for the self-paced reading task consisted of 28 sentences like (8), which were a slightly modified version of the sentences used in Traxler and Pickering (1996).

- (8) a. non-island, implausible

The city that the author wrote regularly about was named for an explorer.

- b. non-island, plausible

The book that the author wrote regularly about was named for an explorer.

c. island, implausible

The city that the author who wrote regularly saw was named for an explorer.

d. island, plausible

The book that the author who wrote regularly saw was named for an explorer.

The implausible and plausible conditions differed only in the filler noun (*city* vs. *book*), which either mismatched or matched with the thematic property of the first verb in the sentence (*wrote*). The non-island and island conditions differed in the number of relative clauses: The non-island condition had only one relative clause (*the city/book that the author wrote regularly about*) such that the verb *wrote* is the first potential gap position, whereas in the island conditions the verb *wrote* is embedded inside another relative clause *the author who wrote regularly*, such that linearly this is still the first verb but grammatically the filler should not be accessible to the verb due to the RC island constraint. Thus, the first verb serves as the critical region for testing the plausibility mismatch effect, and all four conditions include the same adverb in the region after the verb, enabling us to observe a potential spill-over effect. Moreover, all the critical verbs were optionally transitive verbs, such that the sentences in the island conditions end up being grammatical. All target sentences are shown in the Appendix. These 28 sentences were counter-balanced across four lists so that each participant saw only one version of the target items and consequently read 7 tokens from each condition. In addition, 72 fillers of similar length and complexity to the target items were constructed and added to each list.

Acceptability judgment task stimuli. We used a seven-point scale acceptability judgment task to assess participants' knowledge of RC island constraints. We constructed 10 target items like (9):

- (9) a. grammatical sentence: the *wh*-dependency does not cross the RC boundary

The murder case₁ that the law students [_{RC} who learned about the constitution] discussed ____₁ was going to be on the exam.

- b. ungrammatical sentence: the *wh*-dependency crosses the RC boundary

The murder case₁ that the law students [_{RC} who learned about ____₁] discussed ____₁ was going to be on the exam.

The sentences used in this task were modeled after the sentences used in the self-paced reading task, although different lexical items were used so that the participants would not think that they are reading the same sentences as the ones used in the self-paced reading task. In the grammatical condition (9a), the dependency between the filler *the murder case* and the verb *discussed* does not cross the relative clause boundary. It is important to note that the acceptability of (9a) is predicted to be somewhat degraded due to the large processing cost incurred by the presence of more than one temporarily incomplete clause, as revealed by many past empirical studies (e.g., Gibson, 1998; Gibson & Thomas, 1999; Miller & Isard, 1964; Warren & Gibson, 2002; for the effect of number of embedded clauses on acceptability judgments, see Alexopoulou & Keller, 2007).

The ungrammatical counterpart in (9b) was constructed by taking the sentence in (9a) and deleting the object of an obligatorily transitive preposition inside the relative clause, such that the dependency between *the murder case* and the preposition *about* crosses the relative clause boundary. This sentence has a so called parasitic gap configuration (Chomsky, 1982; Culicover, 2001) in that an illicit gap (in this case, the gap inside the relative clause) is followed by a grammatical gap (the complement of *discussed*), but as observed by Engdahl (1983) and Phillips

(2006), the illicit gap inside a relative clause cannot be rendered grammatical by the presence of an additional gap, presumably because the dependency that crosses a strong island like an RC island is strictly prohibited by the grammar (Cinque, 1990).

In order to validate our time-course prediction in the self-paced reading study, it is important to test whether L2 learners have knowledge of the RC island constraint under this parasitic gap configuration. Phillips (2006) demonstrated that the parser postulates a gap inside certain islands if this island-internal gap can be salvaged subsequently as a parasitic gap by an upcoming grammatical gap. If the L2 learners in our present study did not know that gaps inside RC islands cannot be salvaged, then unlike the time-course predication for native speakers discussed above, L2 learners are predicated to demonstrate evidence for active gap creation at the verb inside the RC island. On the other hand, if native speakers and L2 learners both rate the ungrammatical sentence (9b) as significantly more degraded than the grammatical (and yet taxing) sentence (9a), then we can retain the time course prediction discussed above.

The 10 pairs of target items were counter-balanced across two lists, such that each participant only saw one version of a target item and consequently rated 5 grammatical and 5 ungrammatical tokens of (9). These two lists also included 26 fillers with a low to high acceptability range in order to calibrate the participants' use of the acceptability scale.

Procedure

We administered the experiment to the Spanish-English L2 learners in the following order: a) background questionnaire, b) self-paced reading task, c) c-test, and d) paper-and-pencil acceptability judgment task. We decided to conduct the acceptability judgment task after the self-paced reading task so that the sensitive reading time measure would not be affected by having read ungrammatical sentences in the acceptability judgment task that are superficially similar to

the sentences used in the self-paced reading task. The English native speakers only took part in the self-paced reading task and the acceptability judgment task.

The self-paced reading task was implemented on the Linger software developed by Doug Rohde (<http://tedlab.mit.edu/~dr/Linger/>). We used a word-by-word, non-cumulative moving window presentation (Just, Carpenter, & Woolley, 1982). In this design, each sentence initially appears as a series of dashes, and these dashes are replaced by a word from left to right every time the participant presses the space bar. The self-paced reading experiment was preceded by a set of instructions and seven practice items. In order to ensure that the participants were paying attention while reading the sentences, all sentences were followed by yes-no comprehension questions, and feedback was provided if the questions were answered incorrectly.

The acceptability judgment task was administered in a questionnaire format, in which each sentence was accompanied with a seven-point scale (1 was absolutely unacceptable and 7 was perfectly acceptable). Each participant was carefully instructed by the experimenters about the nature of acceptability intuitions and how to use the scale.

Data analysis

For the statistical analyses of the experimental data in this paper, we conduct tests for the native speaker and L2 learner groups separately, since there is no a priori reason to expect that these two groups should be comparable in all respects other than the target phenomena examined in this study. For each statistical test, we report a by-participants analysis ($F1$) and by-items analysis ($F2$) in order to examine the robustness of effects across participants as well as items.

Results

Acceptability judgment task. We report the results from the acceptability judgment task first, as knowledge of island constraints is a necessary premise for interpreting the reading time

data. The acceptability judgment data for the native speaker group and the L2 group are summarized in Figure 1.

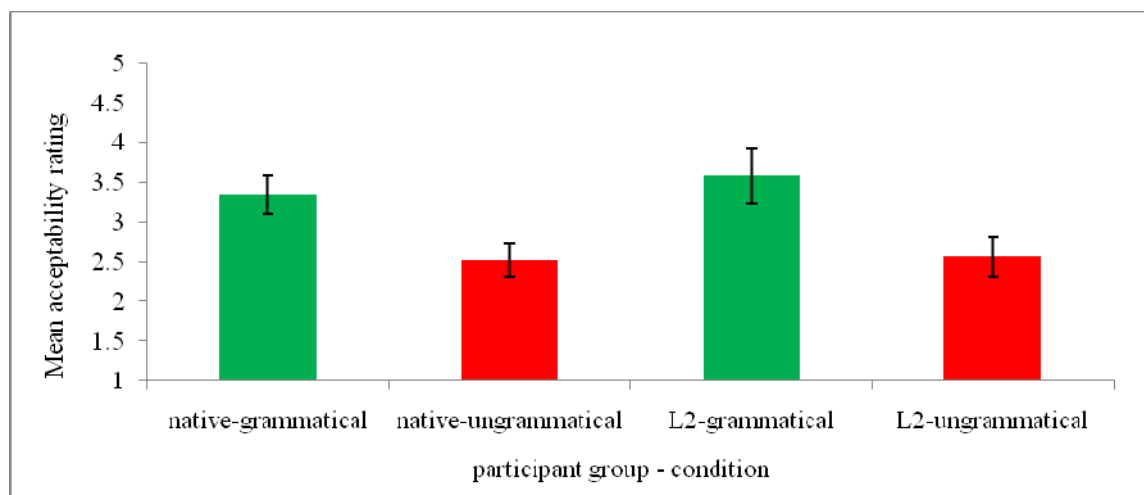


Figure 1. Mean acceptability rating for the native and L2 groups. The error bars indicate the standard error of the mean.

The results from the native speakers confirm the predicted pattern: The mean acceptability rating was 3.35 ($SD = 1.4$) for the grammatical condition and 2.52 ($SD = 1.13$) for the ungrammatical condition. The L2 group showed a similar pattern of contrast: The mean acceptability rating is 3.58 ($SD = 1.72$) for the grammatical condition and 2.57 ($SD = 1.25$) for the ungrammatical condition. The data from the two groups were submitted to separate ANOVAs to test whether the contrasts between the grammatical and ungrammatical conditions were significantly different in each group. For the native speaker group, a repeated measures ANOVA revealed a main effect of grammaticality in by-participants and by-items analyses, $F1(1, 31) = 17.45, p < .05, F2(1, 9) = 9.55, p < .05$, showing that the acceptability rating for the grammatical condition was significantly higher than the rating for the ungrammatical condition. The same pattern of contrast was found in the L2 group, $F1(1, 23) = 12.36, p < .05, F2(1, 9) =$

11.0, $p < .05$, suggesting that the L2 learner group treated the grammatical and ungrammatical conditions in the same way as the native speakers. Taken together, these results indicate that the native speakers and Spanish-English L2 learners have knowledge of the RC island constraint even under this parasitic gap configuration. The present findings are compatible with previous acceptability judgment studies showing that L2 learners whose L1 has overt *wh*-movement are sensitive to island constraints (Martohardjono & Gair, 1993; Schachter, 1990; White, 1988).

One may wonder why the grammatical condition yielded such low ratings in both groups (3.35 out of 7 for the native speakers and 3.58 out of 7 for the non-native speakers, respectively). As discussed in the Method section, this was an expected result based on reports from similar acceptability rating studies that complex grammatical sentences with multiple embeddings routinely receive ratings that are well below ceiling (e.g., Alexopoulou & Keller, 2007). Note also that our ungrammatical condition is similar in overall complexity to the grammatical condition, and in fact the ungrammatical condition is plausibly less complex than the grammatical condition as the former contains one fewer definite NP. Nevertheless, the ungrammatical condition was rated significantly lower than the grammatical condition despite the apparent lower complexity. Therefore, we consider this contrast as a result of the difference in grammaticality.

In summary, the acceptability judgment data revealed that the RC island constraint is observed by both native speakers and non-native speakers in our study. This establishes the premise for the predicted reading time data analysis for the self-paced reading task.

Self-paced reading: Comprehension accuracy. First, mean global accuracy for responses to the comprehension questions for the target and the filler sentences was calculated for the native and non-native speakers. For the native speaker group, the mean accuracy was 90.6% (*SD*

= 5.9) for the target sentences and 86.6% ($SD = 6.4$) for the filler sentences. For the non-native group, the mean accuracy was 92.3% ($SD = 7.1$) for the target sentences and 81.5% ($SD = 5.6$) for the filler sentences. The high comprehension accuracy suggests that both groups of participants were paying attention and carefully reading the sentences for comprehension. Next, we calculated comprehension accuracy for the target sentences in each individual condition. For the native speakers, the mean accuracy was 92.9% ($SD = 10.0$) and 91.5% ($SD = 8.7$) for the non-island implausible and plausible conditions respectively, and 91.5% ($SD = 8.7$) and 86.6% ($SD = 14.5$) for the island implausible and plausible conditions respectively. For the non-native speakers, the mean accuracy was 94.0% ($SD = 23.7$) and 91.7% ($SD = 27.7$) for non-island implausible and plausible conditions respectively, and 92.9% ($SD = 25.8$) and 90.5% ($SD = 29.4$) for island implausible and plausible conditions respectively. The accuracy data were submitted to a repeated measures 2×2 ANOVA with the factors islandhood (whether the critical verb occurs in a non-island or inside an island) and plausibility (whether the filler-verb combination is implausible or plausible) for the native speaker and non-native speaker groups separately. There were no main effects nor significant interactions of any of the factors for the native speakers as well as for non-native speakers (all $F_s < 1$), suggesting that the manipulation of islandhood and plausibility match did not affect the overall comprehension accuracy for the target sentences.

Self-paced reading: Reading time data. Self-paced reading times for the target sentences were examined for each successive region. All trials in which the participant answered the yes-no question incorrectly were excluded. Moreover, trials in which reading time data exceeded three standard deviations from the group mean at each region and in each condition were excluded, which affected less than 1% of trials. The remaining reading time data for each region were submitted to a repeated measures 2×2 ANOVA with the factors islandhood (non-island vs.

island) and plausibility (implausible vs. plausible). The critical regions where a potential plausibility mismatch effect is expected consist of Region 7 (i.e., the verb *wrote* in the example sentence) and the following Region 8 (i.e., the adverb *regularly* in the example sentence) which may reveal a possible spill-over effect, but the non-critical regions are analyzed as well since these regions should exhibit no statistical difference across conditions when they are lexically matched. When the ANOVA results showed a significant interaction in both by-participants and by-items analyses, planned comparisons were conducted to test for the effect of plausibility at each level of the island factor.

The region-by-region mean reading time for the native speaker group is presented in Figure 2.

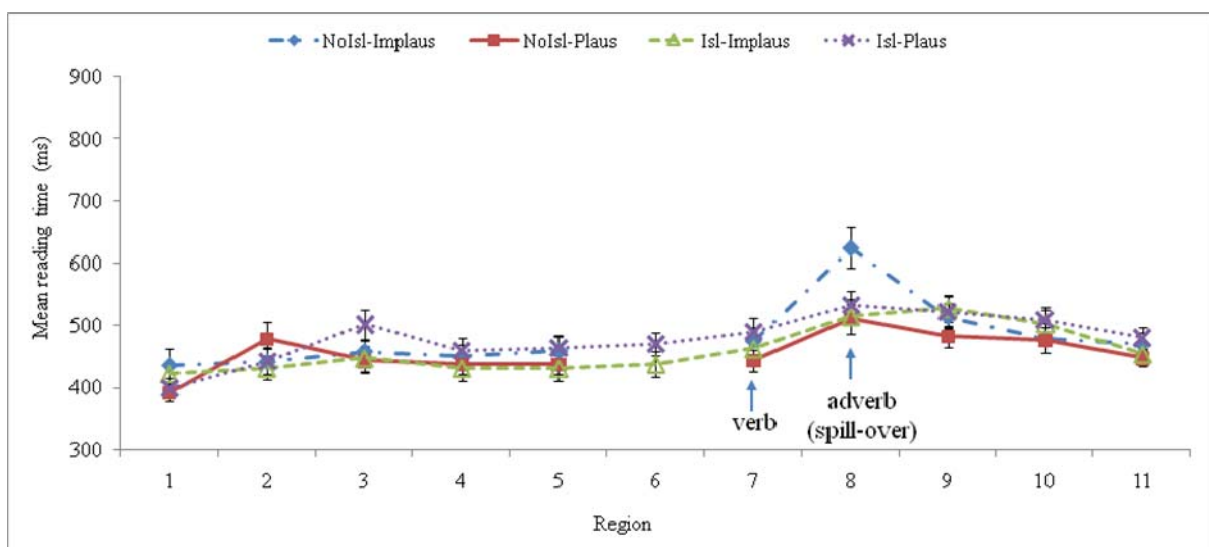


Figure 2. Mean reading time (ms) for the native speaker group. Error bars indicate standard error of the mean.

Sample sentence (words in parentheses appear only in island conditions; words in brackets represent one region):

The₁ city/book₂ that₃ the₄ author₅ (who)₆ wrote₇ regularly₈ about/(saw)₉ was₁₀ [named for an explorer]₁₁.

The statistical analysis of reading time data from each region showed some spurious effects in non-critical regions, but crucially the expected effect in the critical regions (Regions 7 and 8) were larger than the spurious effects.

From Regions 1 to 6, the four conditions were lexically matched and no effects were expected in these regions. However, Region 1 showed a main effect of plausibility in both by-participants and by-items analyses, $F1(1, 31) = 8.56, p < .05, F2(1, 27) = 7.68, p < .05$. In Region 2, a main effect of plausibility was found in by-participants analysis but not in by-items analysis, $F1(1, 31) = 4.59, p < .05, F2(1, 27) = 1.49, p = .23$, and a marginal effect of islandhood in by-items analysis but not in by-participants analysis, $F1(1, 31) = 2.54, p = .121, F2(1, 27) = 3.88, p = .059$. Region 3 similarly revealed a main effect of plausibility only in by-participants analysis, $F1(1, 31) = 4.39, p < .05, F2(1, 27) = 1.67, p = .21$, and a main effect of islandhood in by-items analysis but not in by-participants analysis, $F1(1, 31) = 2.77, p = .1, F2(1, 27) = 6.33, p < .05$, as well as a significant interaction of islandhood and plausibility in both by-participants and by-items analyses, $F1(1, 31) = 7.95, p < .05, F2(1, 27) = 12.03, p < .005$. A pair-wise comparison revealed that there was no significant difference between non-island implausible and plausible conditions (462ms vs. 444ms), $F1(1, 31) = 1.74, p = .176, F2(1, 27) = 1.08, p = .309$, but that the implausible island condition was read faster than the plausible island condition (448ms vs. 501ms), $F1(1, 31) = 10.55, p < .005, F2(1, 27) = 8.74, p < .01$. In Region 4, a significant interaction of islandhood and plausibility was found in by-participants analysis but not in by-items analysis, $F1(1, 31) = 5.25, p < .05, F2(1, 27) = 2.41, p = .13$; and in Region 5,

there was a significant interaction of islandhood and plausibility in by-participants analysis, and a marginally significant interaction in by-items analysis, $F1(1, 31) = 8.56, p < .01, F2(1, 27) = 3.55, p = .07$. We also found a main effect of plausibility in Region 6 in by-participants analysis but not in by-items analysis, $F1(1, 31) = 5.08, p < .05, F2(1, 27) = 2.45, p = .13$.

These spurious effects in Regions 1 through 6 were unexpected since these regions were lexically matched across conditions. However, there are reasons to think that these spurious effects do not undermine data in the critical regions. First, it is known that the first few regions often exhibit greater variability in reading times because participants are sometimes not ready to start reading the sentence. Also, these early regions are quite distant from the critical regions (Regions 7 and 8) and are unlikely to affect the reading time data in the critical regions. Second, the fact that many of these effects did not persist in both by-participants and by-items analyses undermines the robustness of these spurious effects. In fact, these spurious effects are based on much smaller reading time contrasts compared to the reading time contrasts in the critical regions, to which we turn below.

Both Region 7 and Region 8 show the largest effect that replicates a pattern of reading time contrasts observed in Traxler and Pickering (1996), although Region 8 yielded clearer evidence. In both by-participant and by-items analyses, Region 7 showed no main effect of islandhood, $F1(1, 31) = 1.05, p > .1, F2(1, 27) = 1.9, p > .1$, as well as no main effect of plausibility, $F_s < 1$, but there was a significant interaction of islandhood and plausibility, $F1(1, 31) = 6.08, p < .05, F2(1, 27) = 7.68, p < .05$. A planned comparison revealed that the implausible non-island condition was read significantly more slowly than the plausible non-island condition (479ms vs. 444ms), $F1(1, 31) = 4.35, p < .05, F2(1, 27) = 4.4, p < .05$, whereas there was no significant difference between the implausible island and plausible island

conditions (460ms vs. 490ms), $F1(1, 31) = 3.01, p = .093, F2(1, 27) = 3.46, p = .074$. This comparison was marginally significant, but it is unlikely that the marginal significance reflects a plausibility mismatch effect, since the directionality of the difference here is the opposite from the one found in the non-island conditions, i.e., the plausible condition was read more slowly than the implausible condition. The by-participant and by-items analyses of Region 8 revealed a main effect of islandhood, $F1(1, 31) = 10.23, p < .005, F2(1, 27) = 7.22, p < .05$, and a main effect of plausibility, $F1(1, 31) = 4.73, p < .05, F2(1, 27) = 4.8, p < .05$, as well as a significant interaction of islandhood and plausibility, $F1(1, 31) = 13.83, p < .005, F2(1, 27) = 14.51, p < .005$. A planned comparison in this region revealed that the implausible non-island condition was read much more slowly than the plausible non-island condition (623ms vs. 514ms), $F1(1, 31) = 14.24, p < .005, F2(1, 27) = 12.74, p < .005$, but no such difference was found for the island conditions (514ms vs. 530ms), $F_s < 1$.

The fact that the implausible non-island condition was read significantly more slowly than the plausible non-island condition in these critical regions suggests that the parser actively tried to locate a gap at the verb position and consequently experienced a processing difficulty due to the plausibility mismatch. Importantly, however, in the island conditions there was no evidence for active gap filling, suggesting that the island constraint application blocked dependency formation. However, we should be cautious in interpreting the significant interaction in Region 7, since the comparison in the island conditions was marginally significant. This is because the numerical contrasts between island and non-island conditions went in the opposite direction and yielded a cross-over pattern. Note, however, that the analysis of this spill-over region (Region 8) clearly demonstrates evidence for active gap creation in the non-island conditions but not in the island conditions, replicating Traxler and Pickering's findings.

Region 9 contains the actual gap site, and here we found effects that can be attributed to the complexity of the island conditions. In Region 9, the island conditions were read more slowly than the non-island conditions and the difference was marginally significant, $F1(1, 31) = 3.32, p = .078, F2(1, 27) = 2.72, p = .111$. A similar reading time contrast was found in Region 10, $F1(1, 31) = 6.52, p < .05, F2(1, 27) = 2.92, p = .1$, which presumably reflects a spill-over effect from Region 9. These effects seem to reflect the cost of filler-retrieval in island conditions, in which the distance between the filler and the gap is larger than in the non-island conditions.

In Region 11, we found no main effect of island or plausibility, $F_s < 1$, but there was a significant interaction of the two factors, $F1(1, 31) = 15.24, p < .001, F2(1, 27) = 6.08, p < .05$. A planned comparison in Region 11 revealed that the non-island implausible condition was read more slowly than the non-island plausible condition (471ms vs. 447ms) in both by-participants and by-items analyses, $F1(1, 31) = 7.45, p < .05, F2(1, 27) = 4.87, p < .05$, whereas the island implausible condition was read faster than the island plausible condition in by-participants analysis (453ms vs. 482ms), though this effect did not survive in by-items analysis, $F1(1, 31) = 6.0, p < .05, F2(1, 27) = 2.7, p > .1$. The spurious effects in this last region plausibly reflect well-known sentence-final wrap-up effects related to preparation for the comprehension question.

Summarizing so far, in the two critical regions (Regions 7 and 8) we found a pattern of results that replicates Traxler and Pickering's observation: a plausibility mismatch effect occurs in the non-island conditions but not in the island conditions. The presence of unexpected main effects and a significant interaction of islandhood and plausibility in some of the non-critical regions suggests that the reading time data for the native speakers were generally noisy and therefore should be interpreted with caution, but importantly there was no evidence in the critical regions for a plausibility mismatch effect in the island conditions, and the largest effects were

observed in the expected regions. Taken together, it seems reasonable to conclude that the plausibility mismatch effect occurred only in the non-island environment.

Let us turn now to the reading time data for the L2 learner group, which is summarized in Figure 3.

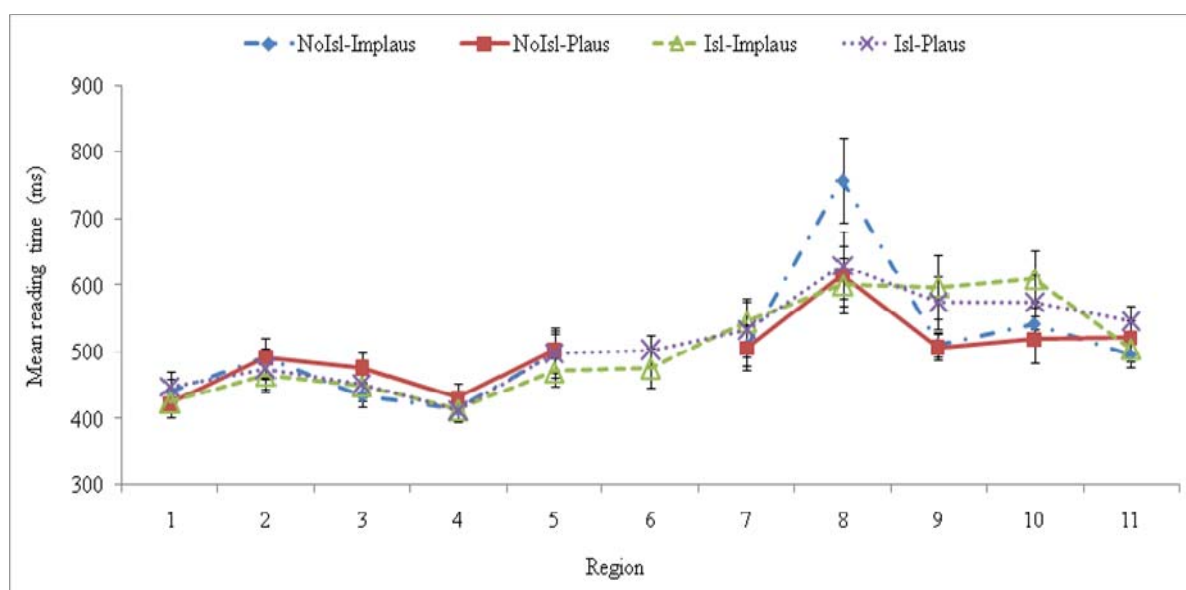


Figure 3. Mean reading time (ms) for the L2 learner group. Error bars indicate standard error of the mean.

Sample sentence (words in parentheses appear only in island conditions; words in brackets represent one region):

The₁ city/book₂ that₃ the₄ author₅ (who)₆ wrote₇ regularly₈ about/(saw)₉ was₁₀ [named for an explorer]₁₁.

We submitted L2 learners' reading time data from each region to a repeated-measures ANOVA with islandhood and plausibility as within-participants factors. Among the non-critical regions, there was no main effect of islandhood, plausibility, nor a significant interaction in Regions 1, 2, 4, 5, and 6. We found a main effect of plausibility in the by-participants analysis of Region 3,

$F1(1, 23) = 4.43, p < .05$, but not in by-items analysis, $F2(1, 27) = 2.34, p = .137$. Since this region is immediately after Region 2 in which plausible and implausible conditions have different lexical items, this may reflect a spill-over effect from Region 2. However, the fact that this did not persist in the by-items analysis suggests that it was not a robust effect.

As in the data for the native speakers, the L2 learners' reading time data showed the largest effects in the critical regions. In Region 7 there was no main effect of islandhood or plausibility, as well as no interaction of the two factors in both by-participants and by-items analyses. In Region 8, on the other hand, there was a main effect of islandhood in both by-participants and by-items analyses, $F1(1, 23) = 6.90, p < .05, F2(1, 27) = 4.52, p < .05$, showing that the non-island conditions were read more slowly than the island conditions. There was also a main effect of plausibility in by-items analysis but not in by-participants analysis, $F1(1, 23) = 2.19, p = .152, F2(1, 27) = 5.11, p < .05$, suggesting that the implausible conditions were read more slowly. We also found a significant interaction of island and plausibility in both by-participants and by-items analyses, $F1(1, 23) = 4.78, p < .05, F2(1, 27) = 9.08, p < .01$. A planned comparison on the island \times plausibility interaction revealed that the non-island implausible condition was read significantly more slowly than the non-island plausible condition (758ms vs. 615ms), $F1(1, 23) = 5.44, p < .05, F2(1, 27) = 16.80, p < .001$, but no such difference was found for the island conditions (600ms vs. 630ms), $F_s < 1$. This pattern of results suggests that the L2 learners actively tried to complete a filler-gap dependency at the critical verb in the non-island conditions only, and this contrast between the non-island and island conditions is exactly the same as the pattern observed for Region 8 in the native speaker group.

We also found a main effect of islandhood in by-participants and by-items analysis of Region 9, $F1(1, 23) = 4.27, p = .05, F2(1, 27) = 9.08, p < .01$, as well as in Region 10, $F1(1, 23)$

= 8.19, $p < .01$, $F2(1, 27) = 8.62$, $p < .01$, suggesting that the island conditions were read more slowly than the non-island conditions. Region 9 is the actual gap site where the filler is retrieved and integrated, and given that the distance of the filler-gap dependency is longer in the island conditions, it seems reasonable that the island conditions should produce slower reading times in this region. Under this interpretation, the fact that Region 10 showed a similar reading time delay in island conditions reflects a spill-over effect from Region 9. Finally, in Region 11 we found a main effect of plausibility in by-participants and by-items analyses, $F1(1, 23) = 4.40$, $p < .01$, $F2(1, 27) = 7.90$, $p < .01$, suggesting that the plausible conditions were read more slowly than the implausible conditions. This wrap-up effect could reflect that the participants were more inclined to reconsider the sentence interpretation in the plausible conditions, but given that this region comes after the critical regions, the data in this region are not informative with respect to the representations that were built during filler-gap dependency processing.

In summary, there were much fewer spurious effects in L2 learners' reading time data, and crucially in Region 8, L2 learners demonstrated evidence for active gap creation in non-island conditions but not in island conditions. This pattern of results replicates the pattern found in our native speaker group as well as in Traxler and Pickering's (1996) original study, suggesting that island constraints successfully blocked ungrammatical long-distance dependency formation in our native and non-native speakers alike.

GENERAL DISCUSSION

The present study tested whether L2 learners can build structural representations with grammatical precision by comparing to what extent advanced Spanish-English L2 learners and English native speakers make use of the relative clause island constraint in constructing a filler-gap dependency. The experiment consisted of an off-line acceptability judgment task to establish

that the L2 learners have the pre-requisite grammatical knowledge, as well as a self-paced reading study to probe the nature of representations constructed during real-time comprehension. Crucially, the self-paced reading study was designed in such a way that the Shallow Structure Hypothesis (SSH) predicted a plausibility mismatch effect in both non-island and island conditions, such that there could be positive evidence for SSH rather than the negative evidence used in previous studies.

The off-line acceptability judgment task examined grammatical knowledge of the RC island constraint under a parasitic gap configuration. Since Phillips (2006) has shown that readers postulate a gap if the possibility of a subsequent licit gap can license a parasitic gap inside a subject island, it was crucial to test whether native speakers as well as non-native speakers can demonstrate knowledge that RC island violations cannot be ameliorated in a parasitic gap configuration. The results revealed that native speakers as well as non-native speakers have the appropriate grammatical knowledge of the RC island constraint, and that the illicit gap inside an RC island cannot be licensed by a later grammatical gap. The finding that L2 learners show knowledge of RC island constraints when their L1 also has an overt *wh*-movement is compatible with previous findings (for a summary, see Belikova & White, 2009), but our results present a novel finding that L2 learners can correctly determine that this illicit gap cannot be remedied by a grammatical gap that appears later in the sentence.

In the on-line self-paced reading measures, the L2 learners and native speakers both demonstrated plausibility mismatch effects when the critical verb was not in an island domain, but importantly, there were no plausibility mismatch effects when the critical verb was embedded inside an RC island. These results demonstrate clear evidence for successful application of the RC island constraint and blocking of ungrammatical long-distance dependency

formation. Our findings suggest that advanced L2 learners not only build structural representations that define an RC island, but also rapidly constrain the active search for a gap location. This further casts doubt on the proposal that L2 learners are unable to build abstract structural representations with grammatical precision.

Effects in the Adverb Region

The evidence for active gap creation in the non-island conditions and the absence of active gap creation in the island conditions was found in the adverb region that immediately followed the critical verb region. Although spill-over effects are extremely common in self-paced reading measures, one may wonder if these results indicate that the participants only constructed filler-gap dependencies based on the information from the adverb itself. In English, adverbs cannot intervene between a verb and an object, and possibly the parser may use this information to realize that the gap must have been present before the adverb. However, we note that the critical verbs that we used in our stimuli were all optionally transitive (e.g., *wrote*), such that even when the adverb is encountered, the gap could be located before the adverb or later in the sentence, e.g., after a PP as in our stimuli. In other words, the information from the adverb still cannot serve as clear bottom-up evidence for where exactly the gap is. Moreover, crucially, even if the adverb played a role in locating the gap, the fact that there was no evidence for gap postulation in the island conditions still indicates that island constraints were respected by native and non-native speakers in their real-time comprehension.

Shallow Structure Hypothesis Revisited

The present findings that L2 learners not only construct detailed structural representations in real-time language processing but also successfully apply island constraints to restrict their active search for a gap location cast doubt on the generality of the shallow structure hypothesis,

and appear to be at odds with the findings from the previous studies that argued for shallow structure building in L2 processing (for a review, see Clahsen & Felser, 2006). Let us now consider why such conflicting findings might have arisen.

There are several reasons to think that some of the evidence for SSH is inconclusive. First, some of the on-line measures that were used to support SSH relied on the lack of target-like reaction time patterns that were predicted based on the assumption that L2 learners have relevant grammatical knowledge, while these studies typically did not have an independent measure to verify the presence of the pre-requisite grammatical competence. As discussed in the introduction, it is possible that these L2 learners were in the process of acquiring target-like grammatical knowledge, and they may have failed to demonstrate target-like behavior in on-line measures precisely for this reason. On-line measures such as reaction times are susceptible to noise and are also likely to reflect a complex interaction of multiple cognitive processes, and it is therefore useful to establish the presence of relevant grammatical knowledge through an off-line measure like an acceptability judgment data as in the present study.

Another reason to question the strength of support for SSH is that some of the arguments relied on ambiguity resolution preferences, i.e., what kind of information the parser prioritizes in the process of selecting structural candidates. For example, Felser, Roberts, Marinis, and Gross (2003) as well as Papadopoulou and Clahsen (2003) report that unlike native speakers, L2 learners show no clear phrase structure-based attachment preferences in ambiguous sentences like *Someone shot the servant of the actress who was on the balcony* where the relative clause can modify either *the servant* or *the actress*, while the L2 learners demonstrated a clear attachment preference based on the lexical semantic property of the preposition *with* in *Someone shot the servant with the actress who was on the balcony*. The asymmetry in the presence of

attachment preferences found in this paradigm has been used to argue for SSH, but alternative interpretations are readily available. It is widely known in the L1 psycholinguistics literature on ambiguity resolution that the parser uses multiple sources of information to select among competing structural candidates (Altmann, 1998; Gibson & Pearlmutter, 1998; Tanenhaus & Trueswell, 1995), and the resolution of relative clause attachment ambiguity investigated by Felser et al. or Papadopoulou and Clahsen is no exception to this, being affected by frequencies (Brysbaert & Mitchell, 1996), prosody (Fodor, 2002), and pragmatic principles (Frazier & Clifton, 1996) to name but a few. Thus, the fact that L2 learners did not show the same phrase structure-based preference as native speakers could simply mean that they ranked non-syntactic information higher in their ambiguity resolution processes (For more discussions of L2 studies on relative clause attachment preferences, see Dussias & Piñar, in press.). In other words, the native vs. non-native contrast in ambiguity resolution studies may reflect “abnormal structural choices” in non-native processing, but this cannot be taken as evidence for SSH which claims that non-native speakers construct “abnormal structural representations.”

The present study specifically focused on L2 learners’ structure generation process rather than the structural selection process. The fact that the L2 learners respected island constraints strongly suggests that the L2 learners were able to build a structural representation with rich syntactic details. In fact, there are other L2 processing studies supporting this view. For example, Rodriguez (2008) examined on-line anaphora resolution in a backward antecedent search in cases where the antecedent comes later than the anaphora, and found that the L2 parser constrains the search domain in accordance with Binding Principle C (Chomsky, 1981), just as was found for native speakers in the studies that this experiment was modeled after (Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2007). This suggests that the L2 parser can construct

abstract structural representations like locality domains or c-command relations between constituents, which challenges SSH because shallow structures are assumed to contain no hierarchical representations (for a similar argument, see also Dekydtspotter, Kim, Kim, Wang, Kim, & Lee, 2008). Taken together, there seems to be stronger evidence for the view that L2 learners can in fact construct structural representations with rich grammatical details.

Are Island Constraints Grammatical Constraints?

There is an on-going debate in the syntax and psycholinguistics literature on the nature of island constraints, that is, whether they are true grammatical constraints that block certain long distance dependencies (McElree & Griffith, 1998; Phillips, 2006; Stowe, 1986; Traxler & Pickering, 1996; Wagers & Phillips, 2009; Yoshida, 2006), or whether island constraints are epiphenomenal and not explicitly represented in a speakers' mind (Deane, 1991; Hawkins, 1999; Hofmeister & Sag, 2009; Kluender, 1998, 2004; Kluender & Kutas, 1993; O'Grady, 2005; Pritchett, 1992). The latter 'reductionist' approach to island constraints attributes island effects to constraints on the parsing mechanism itself. For example, Kluender and his colleagues (Kluender, 1998, 2004; Kluender & Kutas, 1993) suggest that island domains involve complex structural representations and that the cost of processing this domain taxes the memory resources and prevent a retrieval of the filler and subsequent gap filling. On the other hand, Pritchett (1992) attributes the absence of gap filling inside islands to principled structure building procedures that are constrained in such a way that a filler retrieval inside an island domain is prohibited.

The present findings are indeed compatible with both of these accounts. Kluender's resource limitation account of islands would explain the present findings as a reflection of filler-gap association difficulties: the island domains (such as relative clauses) are inherently difficult to process, and for this reason the parser cannot retrieve the filler and associate it with the verb

inside the island, leading to the lack of plausibility mismatch effect in the island conditions. Pritchett's account would explain the present findings as showing that the constraint on the structure building procedures prevented filler-gap associations inside an RC island (see Phillips, 2006, and Sprouse, Wagers & Phillips, 2009, for experimental designs and findings that argue against the reductionist accounts of islands.)

However, it is important to point out that reductionist accounts of islands assume that the parser is capable of building structural representations that trigger a high processing cost or a constraint on parsing procedures. In the case of RC islands, these accounts assume that the parser builds the abstract structural representation of a relative clause, which is responsible for the large processing demand or a parsing constraint that prohibits filler-gap dependency completion. Thus, even if the island constraints turn out to reflect processing factors rather than grammatical knowledge, it does not undermine the logic of the present study in that all of these accounts rely on the parser's ability to build an abstract structural representation like a relative clause, which is precisely what SSH predicts to be unavailable in L2 processing. Therefore, regardless of the nature of island constraints, the present finding presents a clear challenge for the SSH.

CONCLUSION

L2 learners' language comprehension may be less efficient and different in many ways than that of native speakers, but the present study shows that it is not any different with respect to the grammatical details of structural representations built during real-time processing of filler-gap dependencies. The off-line acceptability judgment study confirmed that the native English speakers as well as advanced Spanish-English L2 learners obey island constraints, and the same conclusion was drawn from the on-line self-paced reading study in which both groups showed evidence for active gap filling when the critical verb was not inside an island, but did not show

evidence for active gap filling when the same verb was embedded inside a relative clause island domain. These results strongly suggest that L2 learners can build structural representations that form the basis of an island constraint application, and that they can generally build representations with substantial grammatical precision in real-time processing.

The present finding has implications for L2 processing and acquisition research. First, it casts doubt on views that L2 learners are unable to build abstract structural representations in real-time processing (Clahsen & Felser, 2006). Second, the convergence of off-line and on-line data re-confirms the importance of off-line measures such as acceptability judgment data as a probe for grammatical knowledge in L2 learners. Third, it raises the possibility that some of the differences that were previously found between L1 and L2 processing may be restricted to domains of structure selection and ranking of various sources of information in ambiguity resolution. More generally, the present study highlights the similarity between L1 and L2 processing, lending further support to the view that L1 and L2 linguistic systems are not qualitatively different.

REFERENCES

- Alexopoulou, T., & Keller, F. (2007). Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor. *Language*, 83(1), 110-160.
- Altmann, G. T. M. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences*, 2(4), 146-152.
- Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, 51, 23-54.
- Belikova, A., & White, L. (2009). Evidence for the Fundamental Difference Hypothesis or not? Island constraints revisited. *Studies in Second Language Acquisition*, 31, 199-223.
- Boland, J., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, 34, 774-806.
- Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 49, 664-695.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *Festschrift for Morris Halle* (pp. 232-286). New York: Holt, Rinehart and Winston.
- Chomsky, N. (1977). On wh-movement. In P. W. Culicover, T. Wasow & A. Akmajian (Eds.), *Formal syntax* (pp. 71-132). New York: Academic Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1982). *Some concepts and consequences of the theory of Government and Binding*. Cambridge, MA: MIT Press.

- Chomsky, N. (1986). *Barriers*. Cambridge, MA: MIT Press.
- Cinque, G. (1990). *Types of A'-dependencies*. MIT Press.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3-42.
- Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Karttunen & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives*. Cambridge: Cambridge University Press.
- Culicover, P. W. (2001). Parasitic gaps: A history. In P. W. Culicover & P. M. Postal (Eds.), *Parasitic gaps* (pp. 3–68). Cambridge, MA: MIT Press.
- Dallas, A., & Kaan, E. (2008). Second language processing of filler-gap dependencies by late learners. *Language and Linguistics Compass*, 2, 372–388.
- de Vincenzi, M. (1991). *Syntactic parsing strategies in Italian*. Dordrecht: Kluwer.
- Deane, P. D. (1991). Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics*, 2(1), 1–63.
- Dekydtspotter, L., Kim, B., Kim, H.-j., Wang, Y.-T., Kim, H.-K., & Lee, J. K. (2008). Intermediate traces and anaphora resolution in the processing of English as a second language. In H. Chan, H. Jacob & E. Kapia (Eds.), *Proceedings of the 32nd Annual Boston University Conference on Language Development* (Vol. 1, pp. 84-95). Somerville, MA: Cascadilla Press.
- Dekydtspotter, L., Schwartz, B., & Sprouse, R. (2006). The comparative fallacy in L2 processing research. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings from the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 33–40). Somerville, MA: Cascadilla Press.

- Dussias, P. E., and Piñar, P. (in press). Sentence parsing in L2 learners: Linguistic and experience-based factors. To appear in W. Ritchie & T. Bhatia (Eds.), *The new handbook of second language acquisition*. Amsterdam: Elsevier.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, 23(3), 290.
- Engdahl, E. (1983). Parasitic gaps. *Linguistics and Philosophy*, 6(1), 5–34.
- Engdahl, E. (1997). Relative clause extractions in context. *Working Papers in Scandinavian Syntax*, 60, 51-79.
- Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research*, 23(1), 9.
- Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24, 453-489.
- Fodor, J. D. (2002). Prosodic disambiguation in silent reading. In M. Hirotani (Ed.), *Proceedings of the North East Linguistic Society 32*. Amherst, MA: GLSA Publications.
- Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, 5, 519-559.
- Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4(2), 93-126.
- Frazier, L., & Clifton, C. J. (1996). *Construal*. Cambridge, MA: The MIT Press.
- Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, 18(1), 51-60.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.

- Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences*, 2(7), 262-268.
- Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14, 225-248.
- Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, 7, 55-78.
- Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language*, 75(2), 244-285.
- Hawkins, R., & Hattori, H. (2006). Interpretation of English multiple wh-questions by Japanese speakers: a missing uninterpretable feature account. *Second Language Research*, 22, 269-301.
- Hofmeister, P., & Sag, I. (2009). Cognitive constraints on syntactic islands. An unpublished manuscript, Stanford University and University of California San Diego.
- Huang, J. C. T. (1982). Move WH in a language without WH movement. *The Linguistic Review*, 1, 369-416.
- Inagaki, S. (2001). Motion verbs with goal PPs in the L2 acquisition of English and Japanese. *Studies in Second Language Acquisition*, 23(02), 153-170.
- Jackson, C. N., & Dussias, P. E. (2008). Cross-linguistic differences and their impact on L2 sentence processing. *Bilingualism: Language and Cognition*, 12(01), 65-82.
- Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, 21(2), 121-151.

- Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: Subject and object asymmetries in *wh*-extraction. *Studies in Second Language Acquisition*, 17, 483-516.
- Just, M. A., Carpenter, P., & Wooley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228-238.
- Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, 56(3), 384-409.
- Kluender, R. (1998). On the distinction between strong and weak islands: A processing perspective. In P. Culicover & L. McNally (Eds.), *Syntax and semantics: The limits of syntax* (pp. 241–280). New York: Academic Press.
- Kluender, R. (2004). Are subject islands subject to a processing account. In V. Chand, A. Kelleher, A. Rodriguez, & B. Schmeiser (Eds.), *Proceedings of the 23rd West Coast Conference on Formal Linguistics (WCCFL 23)* (pp. 475–499). Somerville, MA: Cascadilla Press.
- Kluender, R., & Kutas, M. (1993). Subjacency as a Processing Phenomenon. *Language and Cognitive Processes*, 8, 573-633.
- Kush, D., Omaki, A., & Hornstein, N. (2009). Reanalyzing relative clause island effects. A paper presented at Generative Linguistics in the Old World (GLOW) 32. Nantes, France.
- Lasnik, H., & Saito, M. (1992). *Move alpha*. Cambridge, Mass.: MIT Press.
- Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, 33(1), 51-73.

- Li, X. (1998). Adult L2 accessibility to UG: An issue revisited. In S. Flynn, G. Martohardjono & W. O'Neil (Eds.), *The generative study of second language acquisition* (pp. 232-286). Mahwah, NJ: Erlbaum.
- Lieberman, M., Aoshima, S., & Phillips, C. (2006). Nativelike biases in generation of Wh-questions by nonnative speakers of Japanese. *Studies in Second Language Acquisition*, 28(3), 423-448.
- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27, 53-78.
- Martohardjono, G., & Gair, J. (1993). Apparent UG accessibility in second language acquisition: Misapplied principles or principled misapplications. In F. R. Eckman (Ed.), *Confluence: Linguistics, L2 acquisition and speech pathology* (pp. 79-103). Amsterdam: Benjamins.
- McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Learning, Memory*, 24(2), 432-460.
- McKoon, G., Ratcliff, R., & Ward, G. (1994). Testing theories of language processing: An empirical investigation of the on-line lexical decision task. *Learning, Memory*, 20(5), 1219-1228.
- Miller, G. & Isard, S. 1964. Some perceptual consequences of linguistic rules. *Jnl of Verbal Learning & Verbal Behavior* 2: 217-28.
- Montrul, S. (2000). Transitivity alternations in L2 acquisition toward a modular view of transfer. *Studies in Second Language Acquisition*, 22(02), 229-273.
- O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Mahwah, NJ: Erlbaum.

- Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing. *Studies in Second Language Acquisition*, 25(04), 501–528.
- Phillips, C. (2006). The real-time status of island phenomena. *Language*, 82(4), 795–823.
- Phillips, C., & Wagers, M. (2007). Relating structure and time in linguistics and psycholinguistics. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 739–756). Oxford: Oxford University Press.
- Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, 6(3), 229–259.
- Pritchett, B. L. (1992). Parsing with grammar: Islands, heads, and garden paths. In H. Goodluck & M. Rochemont (Eds.), *Island constraints: Theory, acquisition, and processing* (pp. 321–349). Dordrecht: Kluwer Academic Publishers.
- Rizzi, L. (1978). Violations of the WH-Island constraint in Italian. *Journal of Italian Linguistics*, 5, 157–195.
- Rizzi, L. (1990). *Relativized minimality*. Cambridge, MA: MIT Press.
- Rodriguez, G. A. (2008). Second language sentence processing: Is it fundamentally different? Unpublished doctoral dissertation, University of Pittsburgh.
- Ross, J. R. (1967). *Constraints on variables in syntax*. Unpublished doctoral dissertation, MIT, Cambridge.
- Schachter, J. (1990). On the issue of completeness in second language acquisition. *Second Language Research*, 6(2), 93–124.
- Schulz, B. (2006). *Wh-scope marking in English interlanguage grammars: Transfer and processing effects on the second language acquisition of complex wh-questions*. Unpublished doctoral dissertation, University of Hawaii.

- Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Sprouse, J., Wagers, M., & Phillips, C. (2009). Islands and the role of working memory in acceptability judgments. A paper presented at Generative Linguistics in the Old World (GLOW) 32. Nantes, France.
- Stepanov, A. (2007). The end of CED? Minimalism and extraction domains. *Syntax*, 10(1), 80–126.
- Stowe, L. (1986). Evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227–245.
- Sussman, R. S., & Sedivy, J. C. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, 18(2), 143–163.
- Szabolcsi, A., & den Dikken, M. (2003). Islands. In L. Cheng & R. Sybesma (Eds.), *The second glot international state-of-the-article book: the latest in linguistics* (pp. 213–240). Mouton de Gruyter.
- Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of perception and cognition: Speech, language and communication* (pp. 217–262). San Diego: Academic Press.
- Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35(3), 454–475.
- Uriagereka, J. (1999). Multiple Spell-Out. In S. D. Epstein & N. Hornstein (Eds.), *Working Minimalism* (pp. 251–282). Cambridge, MA: MIT Press.

- Wagers, M. W., & Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics*, 45(02), 395–433.
- Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, 85(1), 79-112.
- White, L. (1988). Island effects in second language acquisition. In S. Flynn & W. O'Neil (Eds.), *Linguistic theory in second language acquisition* (pp. 144–172). Dordrecht: Kluwer Academic Publishers.
- White, L., & Juffs, A. (1998). Constraints on wh-movement in two different contexts of nonnative language acquisition: Competence and processing. In S. Flynn, G. Martohardjono & W. O'Neil (Eds.), *The generative study of second language acquisition* (pp. 111-129). Mahwah, NJ: Erlbaum.
- Whong-Barr, M., & Schwartz, B. D. (2002). Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition*, 24(4), 579-616.
- Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, 22, 509-540.
- Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, 9, 71–88.
- Yoshida, M. (2006). *Constraints and mechanisms in long-distance dependency formation*. Unpublished doctoral dissertation, University of Maryland, College Park.

APPENDIX

1a/b. The city/book that the author wrote regularly about was named for an explorer.

1c/d. The city/book that the author who wrote regularly saw was named for an explorer.

2a/b. The injury/officer that the victim called desperately about was treated improperly by doctors.

2c/d. The injury/officer that the victim who called desperately saw was treated improperly by doctors.

3a/b. The dock/rock that the girl threw forcefully at was located near the beach.

3c/d. The dock/rock that the girl who threw forcefully saw was located near the beach.

4a/b. The millionaire/hotel that the architect designed passionately for was nationally well known.

4c/d. The millionaire/hotel that the architect who designed passionately loved was nationally well known.

5a/b. The equipment/consultant that the man phoned hurriedly about was mentioned by the President.

5c/d. The equipment/consultant that the man who phoned hurriedly saw was mentioned by the President.

6a/b. The castle/sign that the tourists read carefully about was photographed by the group.

6c/d. The castle/sign that the tourists who read carefully saw was photographed by the group.

7a/b. The bacteria/researchers that the biologist instructed intensely about turned out to be highly productive.

7c/d. The bacteria/researchers that the biologist who instructed intensely studied turned out to be highly productive.

8a/b. The king/hotel that the custodian cleaned diligently for was ruined by a financial crisis.

8c/d. The king/hotel that the custodian who cleaned diligently served was ruined by a financial crisis.

9a/b. The recording/singer that the instructor taught passionately about was heard throughout the auditorium.

9c/d. The recording/singer that the instructor who taught passionately saw was heard throughout the auditorium.

10a/b. The princess/opera that the musician composed eagerly for was adored by the media.

10c/d. The princess/opera that the musician who composed eagerly admired was adored by the media.

11a/b. The house/letter that the woman wrote cautiously about was inspected by the board.

11c/d. The house/letter that the woman who wrote cautiously saw was inspected by the board.

12a/b. The theories/geniuses that the teacher instructed vigorously about were taught throughout the term.

12c/d. The theories/geniuses that the teacher who instructed vigorously liked were taught throughout the term.

13a/b. The country/general that the soldier killed mercilessly for was destroyed by Mongol military.

13c/d. The country/general that the soldier who killed mercilessly hated was destroyed by Mongol military.

14a/b. The party/lady that the designer dressed elegantly for was thought to be very important.

14c/d. The party/lady that the designer who dressed elegantly enjoyed was thought to be very important.

15a/b. The design/artist that the professor instructed intensely about was discussed in the seminar.

15c/d. The design/artist that the professor who instructed intensely saw was discussed in the seminar.

16a/b. The poster/editor that the manager paid handsomely for was sent to the office.

16c/d. The poster/editor that the manager who paid handsomely saw was sent to the office.

17a/b. The drugs/pupils that the principal asked sternly about were discussed during the meeting.

17c/d. The drugs/pupils that the principal who asked sternly saw were discussed during the meeting.

18a/b. The clock/manuscript that the collector read keenly about was found while shopping for antiques.

18c/d. The clock/manuscript that the collector who read keenly saw was found while shopping for antiques.

19a/b. The client/supper that the cook prepared skillfully for was disliked by the waiters.

19c/d. The client/supper that the cook who prepared skillfully favored was disliked by the waiters.

20a/b. The magazine/clown that the children asked persistently about could not be found anywhere.

20c/d. The magazine/clown that the children who asked persistently saw could not be found anywhere.

21a/b. The match/team that the athlete trained endlessly for was ended by the authorities.

21c/d. The match/team that the athlete who trained endlessly enjoyed was ended by the authorities.

22a/b. The syndicate/executive that the criminal kidnapped cruelly for was absent during the

investigation.

22c/d. The syndicate/executive that the criminal who kidnapped cruelly liked was absent during the investigation.

23a/b. The accident/lawyer that the governor asked cautiously about was seen on the news.

23c/d. The accident/lawyer that the governor who asked cautiously saw was seen on the news.

24a/b. The wall/toy that the boy threw accurately at was painted fire engine red.

24c/d. The wall/toy that the boy who threw accurately noticed was painted fire engine red.

25a/b. The cart/baby that the nanny asked persistently about was seen in the park.

25c/d. The cart/baby that the nanny who asked persistently had was seen in the park.

26a/b. The jewelry/prisoner that the sheriff questioned intensely about was recovered after the robbery.

26c/d. The jewelry/prisoner that the sheriff who questioned intensely watched was recovered after the robbery.

27a/b. The jobs/workers that the instructor taught skillfully about were all in food service.

27c/d. The jobs/workers that the instructor who taught skillfully knew were all in food service.

28a/b. The game/article that the journalist wrote hastily about was discussed at the pub.

28c/d. The game/article that the journalist who wrote hastily saw was discussed at the pub.