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

Akira Omaki
University of Maryland, Université de Genève

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 differ from processing in a native language in a variety of ways, 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 experiments that used stimuli adapted from Traxler and Pickering (1996), both the L2 group and the 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. These results 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 requires the parser to generate grammatically accurate structural representations of the incoming linguistic input. This paper explores the nature of linguistic representations generated by second language (L2) learners during real-time sentence processing, with a particular focus on whether the L2 parser is sensitive to detailed grammatical information.

The issue of the extent to which L2 learners are sensitive to such information relates to the long-standing question about why L2 acquisition appears to be less successful overall as compared to first language (L1) acquisition (Bley-Vroman, 1990, 2009). Based on an extensive review of real-time sentence processing behaviors in child first language learners and adult L2 learners, Clahsen and Felser (2006) proposed the Shallow Structure Hypothesis (henceforth SSH), which states that during real-time language comprehension, L2 learners can only construct shallow structure representations that contain basic argument-predicate relations but lack detailed syntactic information, and therefore their comprehension relies almost exclusively on lexical and semantic information. The SSH is stated to be a general architectural property of the L2 parser; hence the hypothesis predicts that L2 learners can only construct shallow representations regardless of their proficiency levels or how closely related their native and target languages are. This is a very strong claim, but it could potentially provide a nice account for why L2 acquisition is relatively unsuccessful. If L2 learners lack the ability to make use of grammatical information in online sentence processing, then the input that feeds into the language learning mechanism must also be impoverished in its structural details, which may explain why L2 acquisition of grammatical knowledge is not as uniformly successful or efficient as in L1 acquisition.
However, the evidence in support of the SSH is rather inconclusive, and for this reason the present study attempts to further test whether the SSH in its current form is tenable. After reviewing the existing arguments for the SSH and the presence of alternative interpretations for such data, we show how investigating the on-line and off-line use of grammatical constraints such as island constraints on long distance dependency formation can provide a useful testing ground for Clahsen and Felser’s hypothesis. Ultimately, this article challenges the 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 gap creation process (for a similar study that was developed independently of our study and reports compatible findings, see Cunnings, Batterham, Felser and Clahsen, 2009).

A Critical Review of Arguments for the SSH

Clahsen and Felser’s arguments for the SSH have the following form: unlike adult native speakers or children who are learning the target language as their first language, L2 learners a) show lack of sensitivity to syntactic cues in ambiguity resolution, or b) do not demonstrate evidence for constructing abstract representations with grammatical details. We illustrate here that the argument based on the nature of representation is the only relevant type of support for the SSH, but that the existing data provide no conclusive evidence in support of the SSH.

As an example of ambiguity resolution argument, Clahsen, Felser and their colleagues have presented evidence that L2 learners do not demonstrate strong preferences in resolution of relative clause ambiguity resolution in sentences like Someone shot the servant of the actress who was on the balcony (Felser, Roberts, Marinis and Gross, 2003; Papadopoulou and Clahsen, 2003). Clahsen, Felser and colleagues adopt the assumption that native speakers primarily use
phrase-structure based attachment strategies in the absence of lexical-semantic biases (e.g., the use of prepositions like with, as in the servant with the actress), and argued that the lack of preferences in L2 learners follows from their inability to represent grammatical details (for arguments that non-structural factors play critical roles in relative clause attachment preferences even for native speakers, see Brysbaert & Mitchell, 1996; Fodor, 2002; Frazier & Clifton, 1996).

However, 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 fact that L2 learners did not show the same phrase structure-based preference as native speakers could simply mean that non-syntactic information received higher priority in their ambiguity resolution processes. In other words, the native vs. non-native contrast in ambiguity resolution studies may reflect “abnormal structural choices” in non-native processing, but it does not necessarily follow that L2 learners construct “abnormal structural representations,” as proponents of the SSH claim.

Therefore, experiments testing the SSH must be constructed in such a way that on-line, real-time measures such as reading or lexical decision time can directly bear on the nature of the syntactic representations themselves. Here, it is important to note that there are a number of observational and off-line behavioral studies demonstrating rich grammatical knowledge in L2 learners (for a review, see Schwartz, 1998; White, 2003), suggesting that L2 learners are not completely incapable of constructing rich grammatical representations at least when L2 learners are not under time pressure. It is not clear how the SSH accounts for such data, but it is possible that the detailed grammatical representation is constructed only in a second-pass parsing stage but not in first-pass processing, as proposed for native speakers by Townsend and Bever (2001). In this sense, the SSH could be construed as a hypothesis about this first-pass processing.
component, and for this reason, evidence for or against the SSH must come from real-time measures of sentence comprehension.

Some on-line studies have investigated the nature of grammatical representations in 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)). The argument for the SSH in this domain has focused on the representational nature of the gap position, namely, whether it involves a structurally defined abstract ‘trace’ representation. For example, Marinis, Roberts, Felser, and Clahsen (2005) examined processing of a cross-clausal *wh*-dependency (2) 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.

(2)  
   The nurse who<sub>1</sub> the doctor argued t’<sub>1</sub> that the rude patient had angered t<sub>1</sub> 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 (3), 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).

(3) 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.

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. Marinis and colleagues 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).

Although these studies demonstrate behavioral differences between native speakers and L2 learners with respect to filler-gap dependency processing, their findings are open to alternative interpretations. First, it is not clear that the experimental designs used by Marinis and colleagues and by Felser and Roberts 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 and colleagues and in the original study by
Gibson and Warren (2004), although the reading time differences found at the point of filler-retrieval can indicate whether the filler had been previously reactivated or not, this previous reactivation does not necessarily implicate the presence of an abstract representation such as an intermediate trace. It is possible instead that the filler-retrieval and reactivation process is achieved without postulating a trace (for a 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 generally difficult to rely on the cross-modal priming design as a measure of the representations that are constructed during real-time processing.

Second, the findings by Marinis and colleagues and Felser and Roberts are consistent with an interpretation in which L2 learners are using non-target-like parsing procedures while nevertheless building rich structural representations. For example, the lack of reading time evidence for intermediate reactivation in Marinis and colleagues’ study could be the result of a ceiling effect: perhaps 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 (Dekydtspotter, Schwartz, & Sprouse, 2006). The presence of associate priming at both probe positions in Felser and Robert’s study raises the possibility 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 the studies by Marinis and colleagues and Felser and Roberts had different grammatical knowledge than native speakers, but nevertheless
constructed structural representations consistent with their own L2 grammars. Crucially, the hypothesized target-like reading time pattern in Marinis and colleagues’ study relies on the assumption that the participants have target-like grammatical knowledge of \textit{wh}-movement rules and constraints, while Felser and Roberts’ expected pattern of associate priming presumes that the participants had acquired the target-like structural representation of ditransitive constructions, which are known to show complex variations across languages (e.g., Malchukov, Haspelmath, & Comrie, 2010). Neither study, however, provides an independent measure of whether the L2 learners had the relevant grammatical knowledge to demonstrate the expected processing behaviors. It is quite possible that they did not. For example, studies using various off-line measures 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. Li, 1998; Martohardjono & Gair, 1993), and many studies have attested non-target-like argument structure representations in L2 learners arising from cross-linguistic differences in the structure of ditransitive verbs (e.g., Inagaki, 2001; Montrul, 2000; Whong-Barr & Schwartz, 2002). Thus, it is difficult to exclude the possibility that the native vs. non-native processing contrasts in the studies by Marinis and colleagues and Felser and Roberts are not simply due to differences in grammatical knowledge between the two groups.

In summary, the arguments for the SSH are amenable to alternative interpretations. It is possible that non-target-like behaviors of the L2 parser in online processing are related to L2 learners' parsing strategy but not to the nature of their grammatical representations; it is also possible that the L2 learners in these studies simply may have had non-target-like structural representational options. Considering these alternatives, it seems that an ideal test of SSH should have the following components: a) an off-line behavioral task with no time pressure, to examine
whether L2 learners have the pre-requisite knowledge of the relevant rules and constraints, and b) an on-line experimental design that is able to elicit positive time course evidence indicating whether these rules and constraints are deployed with grammatical precision in real-time language processing. Finding both off-line and on-line data that support the presence of rich grammatical details in L2 sentence processing would present a strong challenge against the SSH.

THE PRESENT STUDY: ACTIVE GAP CREATION AND ISLAND CONSTRAINTS

Our study is designed to investigate structural details of filler-gap dependency processing while addressing the methodological concerns discussed above. Specifically, we examine off-line and on-line measures of island constraints on long distance dependency formation, using 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 is capable of eliciting positive reading time evidence for SSH (detailed below).

Psycholinguistic studies on native speakers’ processing of filler-gap dependencies have shown two properties that motivate the use of island constraints in this study. First, the parser shows a strong bias for *active gap creation*, i.e., the parser retrieves and structurally integrates the filler at the earliest potential gap position. For example, an eye-tracking experiment by Traxler and Pickering (1996) examined English-speaking native speakers’ processing of long-distance dependencies in (4).

(4) We like the city / book that the author *wrote* unceasingly and with great dedication about _____ while waiting for a contract.

Traxler and Pickering found a plausibility mismatch effect in (4), 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*), compared to when the filler was a plausible object of the verb (i.e., *the book*). This suggests that as soon as the verb is encountered, the parser immediately creates a gap and analyzes the filler as the object of the verb, despite the fact that the parser could have waited until it encounters the missing argument after the preposition to identify the correct gap position. It has been hypothesized that active gap creation is driven either by a processing principle that requires the parser to complete grammatical dependencies as soon as possible (de Vincenzi, 1991; Frazier, 1987; Pritchett, 1992) or by the need to reduce the cost of retaining the filler in memory (Gibson, 1998). This active gap creation process is robustly attested in L1 parsing in various languages and in a variety of time course measures (e.g., Aoshima, Phillips, & Weinberg, 2004; Crain & Fodor, 1985; Frazier, 1987; Frazier & Clifton, 1989; Garnsey, Tanenhaus, & Chapman, 1989; Lee, 2004; Pickering & Traxler, 2003; Stowe, 1986; Sussman & Sedivy, 2003; see Phillips & Wagers, 2007 for a review of filler-gap dependency processing in native speakers), and similar time-course evidence has been attested to support the existence of active gap creation in L2 sentence processing as well (Jackson & Bobb, 2009; Jackson & Dussias, 2009; Juffs, 2005; Juffs & Harrington, 1995; Williams, Möbius, & Kim, 2001; cf. Williams, 2006; see Dallas & Kaan, 2008 for a review of filler-gap dependency processing in L2 learners).

The second critical property of filler-gap dependency processing relevant for the present study is that despite such a robust preference for immediate gap creation, the parser is sensitive to grammatical constraints on long-distance dependency formation. For example, it has been known since seminal work by Ross (1967) that there are syntactic domains called *islands* that are opaque to syntactic dependency formation (for a review, see Boeckx, 2008; Szabolcsi & den Dikken, 2003). This is illustrated in (5), where an attempt to form a dependency across an island
such as a relative clause (RC) results in an ungrammatical sentence.

(5) *What did the reporter meet the politician [RC who supported ___ at the congress].

In the above-mentioned eye-tracking study by Traxler and Pickering (1996), it was also observed that when the critical verb *wrote* was embedded inside a relative clause island, the plausibility mismatch effect disappeared; i.e., in (6), there was no eye-gaze duration contrast at the verb between *the city* condition and *the book* condition, despite the fact that the same verb was still linearly the first potential gap host after the filler.

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

The absence of active gap creation inside an island for native speakers suggests that the parser applies the island constraint in real-time processing and inhibits ungrammatical long distance dependency formation (for related results, see also McElree & Griffith, 1998; Stowe, 1986; Wagers & Phillips, 2009; Yoshida, 2006). Whether L2 learners also show real-time deployment of island constraints, however, has not been tested.

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 that L2 learners construct for the sentences tested in Marinis and colleagues’ 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] the doctor argued [CP [t'] ] that the rude patient had angered [t ] ]]] 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...*. If L2 learners only construct this kind of shallow representation, then they should not respect the RC island constraint, because there is no relative clause representation in their analysis. The paradigm in (4) and (6) thus presents an ideal testing ground for SSH. If L2 learners actively create gaps while only constructing shallow representations without syntactic details, then the RC island domain cannot be properly represented, and active gap creation (and consequently the plausibility mismatch effect) should be observed in non-island contexts (4) and island contexts (6) 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 (4) but not in (6), replicating the native speaker results from Traxler and Pickering (1996). Thus, unlike the studies by Marinis and colleagues or Felser and Roberts, which presented negative evidence, the current design can potentially elicit positive evidence for SSH in the form of time course data. 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 gap creation 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 (4) and (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 real-time sentence processing. Moreover, this study supplements the on-line self-paced reading study with a separate off-line acceptability judgment study, in which the effect of time pressure and processing limitations can be alleviated to some extent (Chomsky, 1965; Schütze, 1996). This off-line task is used to assess the pre-requisite grammar for our reading time prediction.

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 a review, 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 ($SD = 5.0$).

**Overall 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. An experiment session lasted for approximately 60 minutes. The English native speakers only took part in the self-paced reading task and the acceptability judgment task, and a session lasted approximately 35 minutes.

In the next sections, we will first present the materials, procedures and results for the acceptability judgment task, and then describe those for the self-paced reading task separately.

**Acceptability Judgment Task**

**Materials**

We used a seven-point scale acceptability judgment task to assess participants’ knowledge of RC island constraints. We constructed 10 target items like (8):

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

The murder case$_1$ 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 case$_1$ 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 described below, 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 (8a), 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 (8a) 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 (8b) was constructed by taking the sentence in (8a) and deleting the object of an obligatorily transitive preposition inside the RC, such that the dependency between *the murder case* and the preposition *about* crosses the RC 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 RC) is followed by a grammatical gap (the complement of *discussed*); but as observed by Engdahl (1983) and Phillips (2006), the illicit gap inside a finite RC cannot be rendered grammatical by the presence of an additional gap.

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, contrary to the prediction for native speakers discussed
above, L2 learners are predicted 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 (8b) as significantly more degraded than the grammatical (and yet taxing) sentence (8a), 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 (8). 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.

The acceptability judgment task was administered in a questionnaire format, in which each sentence was accompanied with a seven-point scale (1 being absolutely unacceptable and 7 being 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 all the statistical analyses reported in this paper, we conducted 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 ($F_1$) and by-items analysis ($F_2$) in order to examine the robustness of effects across participants as well as items.

Results

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

[INSERT FIGURE 1 ABOUT HERE]

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; for a review, see Belikova & White, 2009).

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, and we thus consider this contrast to be a result of the difference in grammaticality. In summary, the acceptability judgment data reveal that both native speakers and L2 learners in our study respect the RC island constraint. This establishes the premise for the predicted reading time data analysis for the self-paced reading task.

Self-paced Reading Task

Materials

The materials for the self-paced reading task consisted of 28 sentences like (9), which were a slightly modified version of the sentences used in Traxler and Pickering (1996).

(9)  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 matched or mismatched the selectional restriction property of the first verb in the sentence.
(wrote). The non-island and island conditions differ in the number of relative clauses: The non-island condition has 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. Importantly, 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 are 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 sentence sets 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 were constructed and added to each list.

**Procedure**

The self-paced reading task was implemented on the Linger software developed by Doug Rohde ([http://tedlab.mit.edu/~dr/Linger/](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.
Results

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. The by-condition accuracy data were submitted to a repeated measures 2 \(\times\) 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 or significant interactions of any of the factors for the native speakers or non-native speakers (all \(F_s < 1)\), suggesting that the manipulation of islandhood and plausibility did not affect the comprehension accuracy for the target sentences.

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 \(\times\) 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. The non-critical regions are analyzed as well since these regions should exhibit no statistical difference across conditions given that they are lexically matched. When the critical comparisons showed a significant interaction, planned comparisons were conducted to test for the effect of plausibility at each level of the island factor.

Native speakers’ reading time data

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

The statistical analysis of reading time data revealed some spurious effects in non-critical regions, but crucially the expected effects 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, but the spurious effects with p-values below .05 in both by-participant and by-items analyses are reported below. Region 1 showed a main effect of plausibility, $F(1, 31) = 8.56, p < .05$, $F(1, 27) = 7.68, p < .05$. Region 3 revealed a significant interaction of islandhood and plausibility, $F(1, 31) = 7.95, p < .05$, $F(1, 27) = 12.03, p < .005$. However, 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, 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, $F(1, 31) = 1.05, p > .1$, $F(2, 27) = 1.9, p > .1$, as well as no main effect of plausibility, $F$'s < 1, but there was a significant interaction of islandhood and plausibility, $F(1, 31) = 6.08, p < .05$, $F(2, 27) = 7.68, p < .05$. A planned comparison revealed that the
mean reading time in the implausible non-island condition was reliably slower than in the plausible non-island condition (479ms vs. 444ms), $F_1(1, 31) = 4.35, p < .05, F_2(1, 27) = 4.4, p < .05$, whereas there was no significant difference between the implausible island and plausible island conditions (460ms vs. 490ms), $F_1(1, 31) = 3.01, p = .093, F_2(1, 27) = 3.46, p = .074$. 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. Given that the numerical contrasts between island and non-island conditions went in the opposite direction and yielded a cross-over pattern, we should be cautious in attributing this interaction to the expected result of island constraint application.

The spill-over region presents a much stronger replication of Traxler and Pickering’s results. The by-participant and by-items analyses of Region 8 revealed a main effect of islandhood, $F_1(1, 31) = 10.23, p < .005, F_2(1, 27) = 7.22, p < .05$, and a main effect of plausibility, $F_1(1, 31) = 4.73, p < .05, F_2(1, 27) = 4.8, p < .05$, as well as a significant interaction of islandhood and plausibility, $F_1(1, 31) = 13.83, p < .005, F_2(1, 27) = 14.51, p < .005$. A planned comparison in this region revealed that the reading time in the implausible non-island condition was reliably slower than in the plausible non-island condition (623ms vs. 514ms), $F_1(1, 31) = 14.24, p < .005, F_2(1, 27) = 12.74, p < .005$, 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 creation, suggesting that
the island constraint application blocked dependency formation.

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, $F_1(1, 31) = 15.24, p < .001$, $F_2(1, 27) = 6.08, p < .05$. 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 spurious effects in some of the non-critical regions suggests that the reading time data for the native speakers were slightly 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.

**Non-native speakers’ reading time data**

The reading time data for the L2 learner group are summarized in Figure 3.

L2 learners’ reading time data from each region were submitted 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, $F_1(1, 23) = 4.43, p < .05$, but not in by-items analysis, $F_2(1, 27) = 2.34, p = .137$. The fact that this did not persist in the by-items analysis suggests that it was not a robust effect.

Let us turn to the L2 learners’ reading time data 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, we found exactly the same pattern as native speakers’ data. There was a main effect of islandhood in both by-participants and by-items analyses, $F(1, 23) = 6.90, p < .05$, $F(1, 27) = 4.52, p < .05$, as well as a main effect of plausibility in by-items analysis but not in by-participants analysis, $F(1, 23) = 2.19, p = .152$, $F(1, 27) = 5.11, p < .05$. We also found a significant interaction of island and plausibility in both by-participants and by-items analyses, $F(1, 23) = 4.78, p < .05$, $F(1, 27) = 9.08, p < .01$. A planned comparison on the island × plausibility interaction revealed that the reading time in the non-island implausible condition was reliably slower than the non-island plausible condition (758ms vs. 615ms), $F(1, 23) = 5.44, p < .05$, $F(1, 27) = 16.80, p < .001$, but no such difference was found for the island conditions (600ms vs. 630ms), $Fs < 1$. This pattern of results suggests that the L2 learners actively constructed a gap 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 Region 9, as the reading time in the island conditions were reliably slower than in the non-island conditions, $F(1, 23) = 4.27, p = .05$, $F(1, 27) = 9.08, p < .01$, as well as in Region 10, $F(1, 23) = 8.19, p < .01$, $F(1, 27) = 8.62, p < .01$. 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, as the plausible conditions yielded a slower
reading time the implausible conditions, $F_1(1, 23) = 4.40, p < .01, F_2(1, 27) = 7.90, p < .01$. 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 had 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 Constraints in Second Language Processing

The 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 have 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.

Are Island Constraints Grammatical Constraints?

Our argument against the SSH relies on the real-time application of island constraints, but the nature of these constraints is controversial. There is an on-going debate in the syntax and
psycholinguistics literature on whether island constraints 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 they are epiphenomenal and not explicitly represented in a speakers’ mind (Deane, 1991; Hawkins, 1999; Hofmeister & Sag, 2010; 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. 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 (but see Phillips, 2006, and Sprouse, Wagers & Phillips, submitted, for experimental designs and findings that challenge 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, and it is this abstract structural representation that is responsible for the large processing demand or a parsing constraint that prohibits filler-gap dependency completion inside RCs. 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.

Shallow Structure Hypothesis Revisited

The present study specifically focused on L2 learners’ structure generation process rather than the structure selection process as seen in previous ambiguity resolution studies (Felser et al., 2003; Papadopoulou & Clahsen, 2003). The fact that the L2 learners respected island constraints strongly suggests that they were able to build a structural representation of relative clauses with rich syntactic details.

In fact, an L2 processing study on co-reference restrictions by Rodriguez (2008) lends support to this view. Rodriguez 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 for c-command relations between constituents. This is incompatible with the 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 is stronger evidence for the view that L2 learners are in principle capable of constructing structural representations with rich grammatical details, suggesting that the SSH as a general description of the L2 parser may not be tenable.
On the other hand, the SSH could be maintained in a slightly weaker form, namely that L2 learners might construct shallow structures more often than native speakers do, perhaps in some restricted contexts. This would allow us to retain the intuition behind the SSH that processing factors may play a role in general lack of success in L2 acquisition. For example, it is possible that shallow structures may be more widely observable when the learner’s target language requires grammatical structures and features that are not present in their L1, plausibly because their L1 parser may interfere with parsing of the L2 input (for discussions of this alternative possibility, see Clahsen & Felser, 2006, p. 35; Dussias & Piñar, 2009; Hopp, 2006, 2010). Alternatively, shallow structures may only reflect difficulties in execution of sentence processing procedures. Sentence processing is a complex cognitive task involving lexical access, structure building, semantic composition and discourse integration, and it seems reasonable to think that the parser may attempt to reduce some of the processing burden by adopting less complicated representational options (for a related suggestion for native speakers’ sentence comprehension, see Ferreira & Patson, 2007). Under this view, it is predicted that shallow structures would be adopted less often as the L2 learner's overall proficiency increases. Our Spanish-English L2 learners were fairly advanced learners of English, but less proficient learners may not be able to deploy island constraints in filler-gap dependency processing.

It is also possible that adult L2 learners can use their pragmatic competence to infer the intended message so efficiently that they rely less on precise grammatical structures in their comprehension processes. Consequently, less structural input may be entering their L2 learning mechanism. On the other hand, it has been observed that children are more faithful to their structural analyses and willing to ignore contextual or pragmatic information (e.g., Noveck, 2001; Trueswell, Sekerina, Hill & Logrip, 1999), which might serve to increase the amount of
detailed structural information entering the language learning mechanism. Further investigations of these differences between child L1 processing and adult L2 processing could potentially shed light on the difference in overall success between child L1 acquisition and adult L2 acquisition.

CONCLUSION

L2 learners’ language comprehension may differ in many ways from that of native speakers, but the present study suggests that the L2 parser is not deficient in its representational capacity. The off-line acceptability judgment study confirmed that the native English speakers as well as advanced Spanish-English L2 learners in our study obey island constraints. The same conclusion follows from the on-line self-paced reading study, in which both groups showed evidence for active gap creation when the critical verb was not inside an island, but did not show evidence for active gap creation 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 the RC 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. Second, the convergence of off-line and on-line data reaffirms the importance of off-line measures such as acceptability judgment data as a probe for grammatical knowledge in L2 learners. Third, these results raise 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 support to the view that L1 and L2 linguistic systems are not qualitatively different.
REFERENCES


A. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives*. Cambridge: Cambridge University Press.


Testing, 23(3), 290-325.


Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. Language and Cognitive Processes, 14,


Island Constraints in Second Language Processing


Oxford: Oxford University Press.


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.
Author Note

We would like to thank Theres Grüter, Colin Phillips, Bonnie D. Schwartz, the audience at SLRF 2008, and five anonymous *SSLA* reviewers for their helpful comments and discussions. We also wish to thank Imogen Davidson White, Sherry Warren and Analia Gutierrez for their help with this research project, as well as Alan Mishler for his editorial assistance. This work was supported in part by University of South Carolina Research Opportunity Grant 12530-KA12 awarded to BS.

Address correspondence to: Akira Omaki, Department of Linguistics, 1401 Marie Mount Hall, College Park, MD, 20742. Email: omaki@umd.edu
Figure Captions

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

*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.

*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.