Studies in Second Language Acquisition, 2011, **33**, 563–588.
doi:10.1017/S0272263111000313

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

Akira Omaki
Johns Hopkins University

Barbara Schulz
University of South Carolina

Second-language (L2) sentence 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 (e.g., Clahsen & Felser, 2006). The present study challenges this hypothesis by comparing the extent to which advanced L1 Spanish-L2 English learners and English native speakers make use of the relative clause island constraint in constructing filler-gap dependencies. In offline acceptability judgment and online self-paced reading experiments that used stimuli adapted from Traxler and Pickering (1996), both the L2 group and the native-speaker control group demonstrated clear evidence for application of the relative clause island constraint. These findings suggest that advanced L2 learners not only build abstract structural representations but also rapidly constrain the active

We would like to thank Theres Grüter, Colin Phillips, Bonnie D. Schwartz, the audience at the Second Language Research Forum 2008, and five anonymous *SSLA* reviewers for their helpful comments and discussions. We are grateful to Matt Traxler for sharing his experimental materials. 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 the second author.

Address correspondence to: Akira Omaki, Department of Cognitive Science, Johns Hopkins University, Room 237 Krieger Hall, 3400 North Charles Street, Baltimore, MD 21218; e-mail: omaki@cogsci.jhu.edu.

© Cambridge University Press 2011

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

search for a gap location. These results cast doubt on the proposal that L2 learners are unable to build structural representations with grammatical precision.

Successful language comprehension requires the parser to generate grammatically accurate structural representations of the incoming linguistic input. This study 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 SLA appears to be less successful overall compared to first language (L1) acquisition (Bley-Vroman, 1990, 2009). On the basis of an extensive review of real-time sentence processing behaviors in child L1 and adult L2 learners, Clahsen and Felser (2006) proposed the shallow structure hypothesis (SSH), which states that during real-time language comprehension, L2 learners can only construct shallow representations that contain basic argument-predicate relations but lack detailed syntactic information; 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; thus, the hypothesis predicts that L2 learners can only construct shallow representations regardless of their proficiency levels or how closely related the L1 and L2 are. This is a very strong claim, but it could potentially provide a nice account for why SLA 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 SLA 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 a review of the existing arguments for the SSH and the presence of alternative interpretations for such data, a discussion follows of how investigating the online and offline use of grammatical constraints such as island constraints on long-distance dependency formation can provide a useful testing ground for Clahsen and Felser's (2006) hypothesis. This article ultimately challenges the SSH and presents evidence that both L1 speakers and L2 learners respect island constraints and inhibit ungrammatical long-distance dependency formation, which suggests that L2 learners can build abstract structural representations to guide the parser's active gap creation process (for a similar study that reports compatible findings, see Cunnings, Batterham, Felser, & Clahsen, 2009).

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

A CRITICAL REVIEW OF ARGUMENTS FOR THE SSH

Clahsen and Felser's (2006) arguments for the SSH take the following form: Unlike adult native speakers or children who are learning the target language as their L1, 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. The argument on the basis of the nature of representation, as illustrated in this section, is the only relevant type of support for the SSH, but 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 the resolution of relative clause ambiguity resolution in sentences like *Someone shot the servant of the actress who was on the balcony* (Felser, Roberts, Marinis, & Gross, 2003; Papadopoulou & Clahsen, 2003). Clahsen, Felser, and colleagues adopted 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 nonstructural 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 nonsyntactic information received higher priority in their ambiguity resolution processes. In other words, the native versus nonnative contrast in ambiguity resolution studies may reflect abnormal structural choices in nonnative 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 online, real-time measures such as reading or lexical decision time can directly bear on the nature of the syntactic representations themselves. It is important to note here that there are a number of observational and offline behavioral studies demonstrating rich grammatical knowledge in L2 learners (for a review, see Schwartz, 1998; White, 2003), which suggests that L2 learners are not completely incapable of constructing rich grammatical representations, at least when they 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

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

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 online 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)—to assign a thematic interpretation to the dislocated constituent (called the filler; boldface type 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 crossclausal *wh*-dependency (2) by English native speakers and advanced L1 Chinese, Japanese, German, and Greek learners of English. They found a contrast between the native and nonnative speakers, in that the nonnative speakers showed no clear reading-time evidence for a pregap reactivation of the filler, and they concluded from this finding that the L2 learners failed to postulate an intermediate trace (t') at the clause boundary.

(2) *The nurse who*$_i$ *the doctor argued* t'$_i$ *that the rude patient had angered* t$_i$ *is refusing to work late.*

Felser and Roberts (2007) found a similar native versus nonnative contrast in a crossmodal picture priming study with native speakers of English and advanced L1 Greek-L2 English 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 (pregap probe).

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

In contrast, 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

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

memory, rather than retrieving the filler at the structurally defined gap position. Marinis et al. (2005) as well as Felser and Roberts interpreted these data to indicate that L2 learners do not postulate abstract representations like traces of 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 L1 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 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 L1 speakers. For example, in the experimental design used in Marinis et al. 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, 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 Roberts' study, McKoon, Ratcliff, and Ward (1994) demonstrated that the crossmodal priming effects reported in the L1 psycholinguistics literature are not consistently replicable, which suggests that it is generally difficult to rely on the crossmodal priming design as a measure of the representations that are constructed during real-time processing.

Second, the findings by Marinis et al. (2005) and Felser and Roberts (2007) are consistent with an interpretation in which L2 learners are using nontargetlike 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 Roberts's (2007) 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's 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 et al. (2005) and Felser and Roberts (2007) had different grammatical knowledge than native speakers but, nevertheless, constructed structural

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

representations consistent with their own L2 grammars. It is crucial to note that the hypothesized targetlike reading time pattern in Marinis and colleagues' study relies on the assumption that the participants have targetlike grammatical knowledge of *wh*-movement rules and constraints, whereas Felser and Roberts's expected pattern of associate priming presumes that the participants had acquired the targetlike structural representation of ditransitive constructions, which are known to show complex variation across languages (e.g., Malchukov, Haspelmath, & Comrie, 2010). Neither study, however, provided 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 offline measures have suggested that L2 learners do not always show nativelike sensitivity to locality constraints on long-distance movement (e.g., Hawkins & Hattori, 2006; Schachter, 1990; but see Li, 1998; Martohardjono & Gair, 1993), and many studies have attested nontargetlike argument structure representations in L2 learners arising from crosslinguistic 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 versus nonnative processing contrasts in the studies by Marinis et al. 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 nontargetlike 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 nontargetlike structural representational options. Considering these alternatives, it seems that an ideal test of the SSH should have the following components: (a) an offline behavioral task with no time pressure to examine whether L2 learners have the prerequisite knowledge of the relevant rules and constraints and (b) an online 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 offline and online 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

This study is designed to investigate structural details of filler-gap dependency processing while addressing the methodological concerns

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

discussed here. It specifically examines offline and online measures of island constraints on long-distance dependency formation, using (a) an acceptability judgment task to ascertain that the L2 learners have prerequisite targetlike grammatical knowledge in the relevant domain and (b) an experimental design that is capable of eliciting positive reading-time evidence for SSH.

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; that is, 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 L1 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): 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). Additionally, 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, but see 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

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

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), in which 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 eye-tracking study by Traxler and Pickering (1996), it was also observed that when the critical verb *wrote* was embedded inside a RC island, the plausibility mismatch effect disappeared. For example, 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; Yoshida, 2006). Whether L2 learners also show real-time deployment of island constraints, however, has not been tested.

It is important to consider the prediction of the SSH for this paradigm. For this purpose, it is useful to illustrate the exact representations that Clahsen and Felser (2006) proposed that L2 learners construct for the sentences tested in Marinis et al.'s (2005) study. The representation that Clahsen and Felser proposed 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 RC, *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 RC representation in their analysis. The paradigm

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

in (4) and (6) thus presents an ideal testing ground for the 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 nonisland (4) and island contexts (6) alike. In contrast, if L2 learners turn out to be capable of building RC representations that define a 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 et al. or Felser and Roberts, which presented negative evidence, the current design can potentially elicit positive evidence for the SSH in the form of time-course data. It is also important to note that the application of island constraints is orthogonal to the issue of how L2 learners represent gaps (Felser & Roberts; Marinis et al.). What the island constraints do is restrict the domain in which the parser searches for a gap; 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 L1 Spanish-L2 English learners can construct a structural representation for a RC island and, consequently, constrain their active gap search process in real-time sentence processing. Additionally, it supplements the online self-paced reading study with a separate offline acceptability judgment study, in which the effect of time pressure and processing limitations can be alleviated to some extent (Chomsky, 1965; Schütze, 1996). This offline task is used to assess the prerequisite grammar for the reading-time prediction.

METHOD

Participants

Fifty-six participants were recruited from the University of South Carolina and University of Maryland communities. The groups of participants consisted of 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 $10.00 for their participation.

The L2 learners' overall English proficiency was examined by administering a C-test, which measures general language proficiency on the basis of multiple deletions of parts of words in continuous texts (for a review,

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

see Eckes & Grotjahn, 2006). The L2 learners' average score was 42.5 out of 60 points (*SD* = 10.1; range: 24–58). The C-test was not administered to the native speakers in the current study, but according to Schulz (2006), who reported data from 30 native speakers of English who took the same C-test, the average score was 50–60 (*SD* = 7.7; range: 26–59). Given that the score range of a representative sample of native speakers and that of the L2 learners in this study is roughly equivalent, it is clear that this 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

The experiment was administered to the L1 Spanish-L2 English learners in this order: (a) background questionnaire, (b) self-paced reading task, (c) C-test, and (d) paper-and-pencil acceptability judgment task. The acceptability judgment task was conducted 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 experimental session lasted for approximately 60 min. The English native speakers only took part in the self-paced reading task and the acceptability judgment task, and the session lasted approximately 35 min.

The materials, procedures, and results for the acceptability judgment task will be presented first, followed by a description of those for the self-paced reading task.

ACCEPTABILITY JUDGMENT TASK

Materials

A 7-point scale acceptability judgment task was used to assess participants' knowledge of RC island constraints. Ten target items like (8) were constructed for this task:

(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.*

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

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 were 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 RC 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*); however, 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.

To validate the 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 this study did not know that gaps inside RC islands cannot be salvaged, then, contrary to the prediction for native speakers, L2 learners are predicted to demonstrate evidence for active gap creation at the verb inside the RC island. In contrast, if English L1 speakers and L2 learners both rate the ungrammatical sentence (8b) as significantly more degraded than the grammatical (and yet taxing) sentence (8a), then it is possible to retain the time-course prediction.

The 10 pairs of target items were counterbalanced across two lists so that each participant only saw one version of a target item and, consequently, rated five grammatical and five ungrammatical tokens of (8). These two lists also included 26 fillers with a low to high acceptability range 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 7-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.

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

Data Analysis

For all of the statistical analyses reported in this article, tests for the L1 speaker and L2 learner groups were conducted separately because there was not an 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, a by-participant analysis ($F1$) and by-item analysis ($F2$) is reported 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.

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 was 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-participant and by-item 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, $F1(1, 23) = 12.36$,

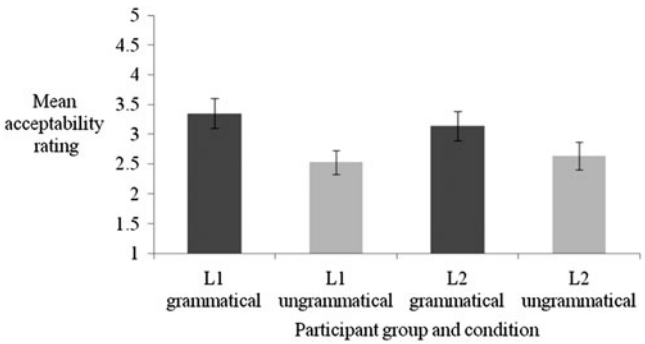


Figure 1. Acceptability judgment data for the native-speaker group and the L2 group.

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

$p < .05$; $F2(1, 9) = 11.0$, $p < .05$, 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 L1 Spanish-L2 English 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).

It may seem surprising that 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 nonnative speakers). 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 the ungrammatical condition here is similar in overall complexity to the grammatical condition, and in fact the ungrammatical condition is plausibly less complex than the grammatical condition, as it contains one fewer definite noun phrase (NP). Nevertheless, the ungrammatical condition was rated significantly lower than the grammatical condition despite the apparent lower complexity, and this contrast is thus considered to be a result of the difference in grammaticality. In summary, the acceptability judgment data reveal that both native speakers and L2 learners in this 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 as in (9), which were a slightly modified version of the sentences used in Traxler and Pickering (1996).

(9) a. Nonisland, implausible
The city that the author wrote regularly about was named for an explorer.
b. Nonisland, 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

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

restriction property of the first verb in the sentence (*wrote*). The nonisland and island conditions differ in the number of RCs: The nonisland condition has only one RC (*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 RC (*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. It is important to notice that all four conditions include the same adverb in the region after the verb, such that this adverb region could be used to observe a potential spill-over effect. Additionally, all of the critical verbs are optionally transitive verbs, so that the sentences in the island conditions end up being grammatical (all target sentences are available on the first author's Web site: http://www.cog.jhu.edu/~omaki/). These 28 sentence sets were counterbalanced across four lists so that each participant saw only one version of the target items and, consequently, read seven tokens from each condition. Additionally, 72 fillers of similar length and complexity were constructed and added to each list.

Procedure

The self-paced reading task was implemented on Linger software (Rohde, 2003), and a word-by-word, noncumulative moving window presentation was used (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. 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 nonnative 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 nonnative group, the mean accuracy was 92.3% (*SD* = 7.1) for the target sentences

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

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 × 2 ANOVA with the factors Islandhood (i.e., whether the critical verb occurs in a nonisland or inside an island) and Plausibility (whether the filler-verb combination is implausible or plausible) for the native-speaker and nonnative-speaker groups separately. There were no main effects or significant interactions of any of the factors for the native speakers or nonnative speakers (all *F*s < 1), which suggests 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 × 2 ANOVA with the factors islandhood (nonisland vs. island) and plausibility (implausible vs. plausible). The critical regions in which 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 noncritical regions are analyzed as well because 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.

L1 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 noncritical regions, but, crucially, the expected effects in the critical regions (Regions 7 & 8) were larger than the spurious effects.

From Region 1 to Region 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-item analyses are reported below. Region 1 showed a main effect of Plausibility, $F1(1, 31) = 8.56$, $p < .05$; $F2(1, 27) = 7.68$, $p < .05$. Region 3 revealed a significant

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

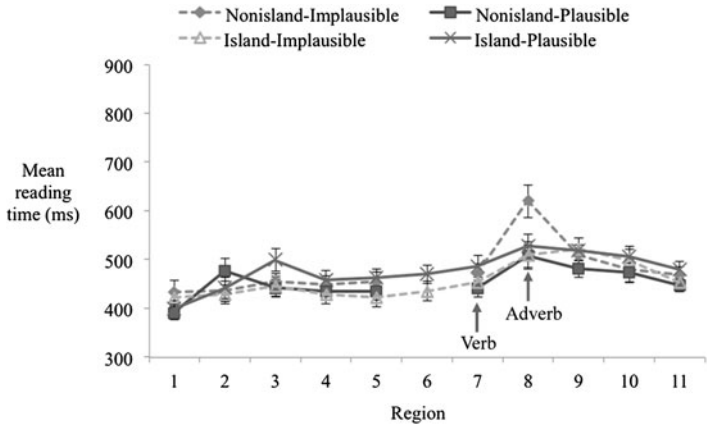


Figure 2. Region-by-region mean reading time for the native speakers.

interaction of Islandhood and Plausibility, $F1(1, 31) = 7.95$, $p < .05$; $F2(1, 27) = 12.03$, $p < .005$. However, these early regions are quite distant from the critical regions (Region 7 & Region 8) and are unlikely to affect the reading-time data in the critical regions.

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-item 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, Fs < 1: however, 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 mean reading time in the implausible nonisland condition was reliably slower than in the plausible nonisland condition (479 ms vs. 444 ms), $F1(1, 31) = 4.35$, $p < .05$; $F2(1, 27) = 4.4$, $p < .05$, whereas there was no significant difference between the implausible island and plausible island conditions (460 ms vs. 490 ms), $F1(1, 31) = 3.01$, $p = .093$; $F2(1, 27) = 3.46$, $p = .074$. This comparison was marginally significant, but it is unlikely that the marginal significance reflects a plausibility mismatch effect because the directionality of the difference here is the opposite from that found in the nonisland conditions; that is, the plausible condition was read more slowly than the implausible condition. Given that the numerical contrasts between island and nonisland conditions went in the opposite direction and yielded a crossover pattern, caution should be exercised 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 (1996) results. The by-participant and by-item analyses of

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

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 reading time in the implausible nonisland condition was reliably slower than in the plausible nonisland condition (623 ms vs. 514 ms), $F1(1, 31) = 14.24$, $p < .005$; $F2(1, 27) = 12.74$, $p < .005$, but no such difference was found for the island conditions (514 ms vs. 530 ms), Fs < 1. The fact that the implausible nonisland condition was read significantly more slowly than the plausible nonisland 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. However, it is important to notice that in the island conditions, there was no evidence for active gap creation, suggesting that the island constraint application blocked dependency formation.

In Region 11, no main effect of island or plausibility was found, Fs < 1, but there was a significant interaction between the two factors, $F1(1, 31) = 15.24$, $p < .001$; $F2(1, 27) = 6.08$, $p < .05$. 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, the two critical regions (Region 7 and Region 8) demonstrated a pattern of results that replicates Traxler and Pickering's (1996) observation: A plausibility mismatch effect occurs in the nonisland conditions but not in the island conditions. The presence of spurious effects in some of the noncritical regions suggests that the reading-time data for the native speakers were slightly noisy and, therefore, should be interpreted with caution; however, more 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 nonisland environment.

Nonnative 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 noncritical regions, there was no main effect of Islandhood and Plausibility, nor a significant interaction in Regions 1, 2, 4, 5, and 6. A main effect of plausibility was found in the by-participant analysis of Region 3, $F1(1, 23) = 4.43$, $p < .05$, but not

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

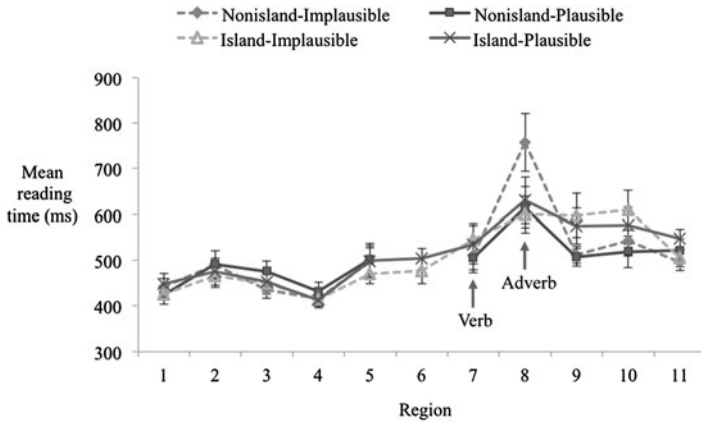


Figure 3. Reading-time data for the L2 learner group.

in by-item analysis, $F2(1, 27) = 2.34$, $p = .137$. The fact that this did not persist in the by-item analysis suggests that it was not a robust effect.

As for the L2 learners' reading-time data in the critical regions, there was no main effect of Islandhood or Plausibility in Region 7 as well as no interaction of the two factors in both by-participant and by-item analyses. In contrast, the exact same pattern as native speakers' data was found in Region 8. There was a main effect of Islandhood in both by-participant and by-item analyses, $F1(1, 23) = 6.90$, $p < .05$; $F2(1, 27) = 4.52$, $p < .05$, as well as a main effect of Plausibility in by-item analysis but not in by-participant analysis, $F1(1, 23) = 2.19$, $p = .152$; $F2(1, 27) = 5.11$, $p < .05$. The analyses also yielded a significant interaction of Island and Plausibility in both by-participant and by-item analyses, $F1(1, 23) = 4.78$, $p < .05$; $F2(1, 27) = 9.08$, $p < .01$. A planned comparison on the Island × Plausibility interaction revealed that the reading time in the nonisland implausible condition was reliably slower than the nonisland plausible condition (758 ms vs. 615 ms), $F1(1, 23) = 5.44$, $p < .05$; $F2(1, 27) = 16.80$, $p < .001$, but no such difference was found for the island conditions (600 ms vs. 630 ms), Fs < 1. This pattern of results suggests that the L2 learners actively constructed a gap at the critical verb in the nonisland conditions only, and this contrast between the nonisland and island conditions is exactly the same as the pattern observed for Region 8 in the native-speaker group.

There was also a main effect of Islandhood found in Region 9, as the reading time in the island conditions were reliably slower than in the nonisland conditions, $F1(1, 23) = 4.27$, $p = .05$; $F2(1, 27) = 9.08$, $p < .01$, as well as in Region 10, $F1(1, 23) = 8.19$, $p < .01$; $F2(1, 27) = 8.62$, $p < .01$. 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

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

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, a main effect of plausibility in by-participant and by-item analyses was found in Region 11, as the plausible conditions yielded a slower reading time than the implausible conditions, $F1(1, 23) = 4.40$, $p < .01$; $F2(1, 27) = 7.90$, $p < .01$. This wrap-up effect could reflect that the participants were more inclined to reconsider the sentence interpretation in the plausible conditions; however, 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 sum, 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 nonisland conditions but not in island conditions. This pattern of results replicates the pattern found in the 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 the native and nonnative speakers alike.

GENERAL DISCUSSION

The present study tested whether L2 learners can build structural representations with grammatical precision by comparing to what extent advanced L1 Spanish–L2 English learners and English native speakers make use of the RC island constraint in constructing a filler-gap dependency. The experiment consisted of an offline acceptability judgment task to establish that the L2 learners had the prerequisite 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 SSH predicted a plausibility mismatch effect in both nonisland and island conditions such that there could be positive evidence for the SSH rather than the negative evidence used in previous studies.

The offline acceptability judgment task examined grammatical knowledge of the RC island constraint under a parasitic gap configuration. Because 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 as well as nonnative speakers have knowledge that RC island violations cannot be ameliorated in a parasitic gap configuration. The results revealed that native as well as nonnative speakers have the appropriate grammatical knowledge

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

of the RC island constraint and that the illicit gap inside a 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). However, these 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 online 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 a RC island. These results demonstrate clear evidence for successful application of the RC island constraint and blocking of ungrammatical long-distance dependency formation. These findings suggest that advanced L2 learners not only build structural representations that define a RC island but also rapidly constrain the active search for a gap location. This casts further doubt on the proposal that L2 learners are unable to build abstract structural representations with grammatical precision.

ARE ISLAND CONSTRAINTS GRAMMATICAL CONSTRAINTS?

The argument against the SSH relies on the real-time application of island constraints, but the nature of these constraints is controversial. There is an ongoing debate in the syntax and psycholinguistics literature on whether island constraints are true grammatical constraints that block certain long-distance dependencies (e.g., McElree & Griffith, 1998; Phillips, 2006; Stowe, 1986; Traxler & Pickering, 1996; Yoshida, 2006) or whether they are epiphenomenal and not explicitly represented in a speakers' mind (e.g., Deane, 1991; Hawkins, 1999; Hofmeister & Sag, 2010; Kluender, 1998, 2004; Kluender & Kutas, 1993; O'Grady, 2005; Pritchett, 1992). This reductionist approach to island constraints attributes island effects to constraints on the parsing mechanism itself. For example, Kluender and his colleagues (e.g., Kluender, 1998, 2004; Kluender & Kutas) suggested that island domains involve complex structural representations and that the cost of processing this domain taxes the memory resources and prevents a retrieval of the filler and subsequent gap filling. The present findings are indeed compatible with both of these accounts. The resource limitation account of islands would explain the present findings as a reflection of filler-gap association difficulties: The island domains (such as RCs) are inherently difficult to process, and for this reason, the parser cannot retrieve the filler and associate it with the verb inside the

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

island, leading to the lack of plausibility mismatch effect in the island conditions (but see Phillips, 2006; and Sprouse, Wagers, & Phillips, in press, 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 RC, 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 RC, which is precisely what the 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.

THE SSH 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 (e.g., 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 RCs with rich syntactic details.

In fact, a L2 processing study on coreference restrictions by Rodriguez (2008) lends support to this view. Rodriguez examined online anaphora resolution in a backward antecedent search in cases for which 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 after which this experiment was modeled (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 et al., 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.

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

However, 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 be consistent with the intuition behind the SSH that processing factors may play a role in general lack of success in SLA. For example, it is possible that shallow structures may be more widely observable when the learner's L2 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; Dussias & Piñar, 2009; Hopp, 2007, 2010). Alternatively, shallow structures may only reflect difficulties in execution of sentence processing procedures. Sentence processing is a complex cognitive task that involves lexical access, structure building, semantic composition, and discourse integration. 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. The L1 Spanish–L2 English learners in this study were fairly advanced learners of English; 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. As a consequence, less structural input may be entering their L2 learning mechanism. However, 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 SLA.

CONCLUSION

Second language learners' language comprehension may differ in many ways from that of L1 speakers, but the present study suggests that the L2 parser is not deficient in its representational capacity. The offline acceptability judgment study confirmed that the native English speakers as well as advanced L1 Spanish–L2 English learners in this study obey island constraints. The same conclusion follows from the online

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

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 RC 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 offline and online data reaffirms the importance of offline 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.

(*Received 3 September 2010*)

REFERENCES

Alexopoulou, T., & Keller, F. (2007). Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor. *Language*, *83*, 110–160.

Altmann, G. T. M. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences*, *2*, 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.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, *20*, 3–49.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition*, *31*, 175–198.

Boeckx, C. (2008). Islands. *Language and Linguistics Compass*, *2*, 151–167.

Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence processing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology*, *49*, 664–695.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT 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.

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. Kartunnen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 94–128). New York: 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.

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

Cunnings, I., Batterham, C., Felser, C., & Clahsen, H. (2009). Constraints on L2 learners' processing of wh-dependencies. Evidence from eye movements. *Essex Research Reports in Linguistics*, *58*, 1–24.
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*, 1–63.
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 Boston University conference on language development 32* (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 of the generative approaches to second language acquisition conference 2006* (pp. 33–40). Somerville, MA: Cascadilla Press.
Dussias, P., & Piñar, P. (2009). Sentence parsing in L2 learners: Linguistic and experienced-based factors. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 296–318). Bingley, UK: Emerald Group.
Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*, 290–325.
Engdahl, E. (1983). Parasitic gaps. *Linguistics and Philosophy*, *6*, 5–34.
Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research*, *23*, 9–36.
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.
Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistic Compass*, *1*, 71–83.
Fodor, J. D. (2002). Prosodic disambiguation in silent reading. In M. Hirotani (Ed.), *Proceedings of the North East Linguistic Society 32* (pp. 113–132). Amherst, MA: GLSA.
Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, *5*, 519–559.
Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, *4*, 93–126.
Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: 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*, 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*, 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*, 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. (2010). Cognitive constraints on syntactic islands. *Language*, *86*, 366– 415.
Hopp, H. (2007). *Ultimate attainment at the interfaces in second language acquisition: Grammar and processing*. Unpublished doctoral dissertation, University of Groningen, The Netherlands.
Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*, 901–931.

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

Inagaki, S. (2001). Motion verbs with goal PPs in the L2 acquisition of English and Japanese. *Studies in Second Language Acquisition*, *23*, 153–170.
Jackson, C. N., & Bobb, S. C. (2009). The processing and comprehension of *wh*-questions among second language speakers of German. *Applied Psycholinguistics*, *30*, 603–636.
Jackson, C. N., & Dussias, P. E. (2009). Cross-linguistic differences and their impact on L2 sentence processing. *Bilingualism: Language and Cognition*, *12*, 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*, 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*, 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). San Diego, CA: 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 west coast conference on formal linguistics 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.
Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, *33*, 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.
Malchukov, A., Haspelmath, M., & Comrie, B. (Eds.) (2010). *Studies in ditransitive constructions: A comparative handbook*. Berlin: Mouton de Gruyter.
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. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 432–460.
McKoon, G., Ratcliff, R., & Ward, G. (1994). Testing theories of language processing: An empirical investigation of the on-line lexical decision task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 1219–1228.
Miller, G., & Isard, S. (1964). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, *2*, 217–228.
Montrul, S. (2000). Transitivity alternations in L2 acquisition toward a modular view of transfer. *Studies in Second Language Acquisition*, *22*, 229–273.
Noveck, I. A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition*, *78*, 165–188.
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*, 501–528.
Phillips, C. (2006). The real-time status of island phenomena. *Language*, *82*, 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*, 229–259.

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press

Pickering, M., & Traxler, M. (2003). Evidence against the use of subcategorization frequency in the processing of unbounded dependencies. *Language and Cognitive Processes*, *18*, 469–503.
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.
Rodríguez, G. A. (2008). *Second language sentence processing: Is it fundamentally different?* Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, PA.
Rohde, D. (2003). Linger software [Computer software]. Retrieved August 1, 2007, from http://tedlab.mit.edu/~dr/Linger/
Ross, J. R. (1967). *Constraints on variables in syntax*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge.
Schachter, J. (1990). On the issue of completeness in second language acquisition. *Second Language Research*, *6*, 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 Hawai'i, Honolulu.
Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
Schwartz, B. D. (1998). The second language instinct. *Lingua*, *106*, 133–160.
Sprouse, J., Wagers, M., & Phillips, C. (in press). A test of the relation between working memory and syntactic island effects. *Language*.
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*, 143–163.
Szabolcsi, A., & den Dikken, M. (2003). Islands. In L. Cheng & R. Sybesma (Eds.), *The second glot international state-of-the-article book* (pp. 213–240). Berlin: 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, CA: 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*, 454–475.
Trueswell, J. C., Sekerina, I., Hill, N. H., & Logrip, L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134.
Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*, 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.
White, L. (2003). *Second language acquisition and Universal Grammar*. New York: Cambridge University Press.
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*, 579–616.
Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, *9*, 71–88.
Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English *wh*- questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, *22*, 509–540.
Yoshida, M. (2006). *Constraints and mechanisms in long-distance dependency formation*. Unpublished doctoral dissertation, University of Maryland, College Park.

https://doi.org/10.1017/S0272263111000313 Published online by Cambridge University Press