

Essex Research Reports in Linguistics

Volume 58

Number 1

16 Sept, 2009

Ian Cunnings

Claire Batterham

Claudia Felser

Harald Clahsen

Constraints on L2 learners' processing of
wh-dependencies: Evidence from eye movements

Dept. of Language and Linguistics,
University of Essex,
Wivenhoe Park,
Colchester, Essex, UK,
CO4 3SQ

<http://www.essex.ac.uk/linguistics/publications/errl/>

Essex Research Reports in Linguistics present ongoing research activities of the members of the Department of Language and Linguistics.

The main purpose of these reports is to provide a quick publication outlet. They have 'pre-publication status', and most will subsequently appear in revised form as research articles in professional journals or in edited books.

Copyright remains with the author(s) of the reports. Comments are welcome: please communicate directly with the authors.

If you have technical problems downloading a paper, or for further information about these reports, please contact the editor:

Doug Arnold: doug@essex.ac.uk.

Citation Information:

Ian Cunnings, Claire Batterham, Claudia Felser and Harald Clahsen. 'Constraints on L2 learners' processing of wh-dependencies: Evidence from eye movements', *Essex Research Reports in Linguistics*, Vol. 58.1. Dept. of Language and Linguistics, University of Essex, Colchester, UK, Sept, 2009.

<http://www.essex.ac.uk/linguistics/publications/errl/errl58-1.pdf>

Constraints on L2 learners' processing of *wh*-dependencies: Evidence from eye movements

Ian Cunnings, Claire Batterham, Claudia Felser & Harald Clahsen
University of Essex

Abstract

Using the eye movement monitoring technique, the present study examines whether *wh*-dependency formation is sensitive to island constraints in second language (L2) sentence comprehension, and whether the presence of an intervening 'pseudo-gap' or relative clause island has any effects on learners' ability to ultimately resolve long *wh*-dependencies. Participants included proficient learners of L2 English from typologically different language backgrounds (German, Chinese), as well as a group of native English-speaking controls. Our results indicate that both the learners and the native speakers were sensitive to relative clause islands during processing, irrespective of typological differences between the learners' L1s, but that the learners had more difficulty than native speakers linking distant *wh*-fillers to their lexical subcategorizers during processing. We provide a unified processing-based account for our findings.

1 Introduction

Unbounded dependencies pose a challenge for real-time language comprehension in that a fronted constituent (or 'filler') needs to be stored temporarily in working memory until it can be linked to its lexical licenser or corresponding 'gap' further downstream, as indicated in (1) below.

- (1) *Which book* did the incredibly famous author claim she wrote __ in just a week?

Psycholinguistic evidence shows that the parser will normally strive to keep filler-gap dependencies as short as possible by trying to link a filler to the earliest potential subcategorizer that it encounters (Frazier & Clifton, 1989). There is evidence for the use of an 'active filler strategy' also in non-native sentence processing (Williams, Möbius, & Kim, 2001; Williams, 2006), although L2 learners' processing of *wh*-dependencies has been found to be non-nativelike in other respects, and even at higher levels of L2 proficiency.

The results from L2 reading-time studies by Marinis, Roberts, Felser, and Clahsen (2005) and Felser & Roberts (2007), for example, indicate that learners of English from both *wh*-movement and *wh*-in situ backgrounds fail to link *wh*-fillers to purely structurally defined gaps (i.e. gaps that are non-adjacent to the filler's lexical subcategorizer). Marinis et al. examined the processing of sentences involving successive-cyclic *wh*-movement such as (2) below by proficient L2 learners of English from typologically different language backgrounds.

- (2) The nurse [_{CP} *who* the doctor argued [_{CP} ____ #1 that the rude patient had angered ____ #2]] is refusing to work late.

The analysis of participants' segment-by-segment reading times showed that filler integration at the embedded verb *angered* (i.e., at gap #2) was facilitated by the availability of an intermediate structural gap (= #1) at the left edge of the embedded CP, compared to a gap-less control condition of the same length, for English native speakers only. None of the L1 Chinese, Japanese, German or Greek learner groups showed any reading-time evidence for postulating intermediate structural gaps, however. Marinis et al.'s findings were recently replicated by Pliatsikas and Marinis (2009), who found that even advanced Greek-speaking learners with many years of naturalistic exposure to English showed no evidence of postulating intermediate structural gaps in L2 processing.

Using the cross-modal priming paradigm, Felser and Roberts (2007) investigated antecedent reactivation at indirect object gaps in sentences such as (3) below, in proficient Greek-speaking learners of English.

- (3) Bob loved the monkey *to which* the fat squirrel showed his excellent #1 new trick ____ #2 in the playground last month.

While listening to sentences like the above, only the native speaker controls showed evidence of mentally reactivating the referent of the *wh*-pronoun's antecedent, *the monkey*, at the indirect object gap (= #2) but not at a pre-gap control position 500 ms earlier (= #1). The learners, on the other hand, did not show any position-specific antecedent reactivation effects.

Together with other psycholinguistic and neurolinguistic evidence indicating L1/L2 differences in the domain of grammatical processing, these findings led Clahsen and Felser (2006) to propose the *Shallow Structure Hypothesis* (SSH) for L2 processing according to which non-native speakers have difficulty building complex grammatical representations during real time L2 comprehension and production. The SSH claims that L2 processing is more semantics-driven, and less guided by morphosyntactic or other grammatical cues to interpretation, than native language processing. For the L2 processing of filler-gap dependencies, this means that learners should be able to associate a filler with its lexical subcategorizer but not necessarily link it to gaps whose presence is contingent on the parser's building abstract hierarchical representations (compare Felser & Roberts, 2007; Marinis et al., 2005; Pliatsikas & Marinis, 2009).

While there is potentially no limit as to the linear distance between a *wh*-filler and its (ultimate) associated gap, in *wh*-movement languages like English, unbounded dependencies are subject to subjacency or 'island' constraints that prohibit dependency formation in certain syntactic environments (Ross, 1967). *Wh*-extraction from relative clauses (RCs) as in (4b) below, for example, is normally ruled out, rendering RCs syntactic islands.

- (4) a. The author [_{RC} who wrote a book on internet crime] gave an interview yesterday.
- b. *Which book did the author [_{RC} who wrote ____] give an interview yesterday?

In the generative tradition, island effects have been assumed to reflect universal grammatical constraints such as the subjacency condition (Chomsky, 1973), which prohibits movement across more than one 'bounding node' at a time. A number of observations call into question the view that island constraints are purely grammatical phenomena, however. These include the fact that acceptability ratings often vary, and exceptions such as (5) below (from Chung & McCloskey, 1983) are well attested.

- (5) a. We have a visitor who there's no one who's willing to host.
- b. This is a paper that we really need to find someone who understands.

The use of definite noun phrases and finite verb forms, for instance, can increase the perceived severity of island violations, compared to indefinites or covert pronouns and non-finite forms (Ross, 1967). Compare (6a) and (6b) below, the latter of which is quite acceptable (examples adapted from Kluender, 2004).

- (6) a. That's the campaign that I finally thought of the aide [_{RC} who could
spearhead __]
b. That's the campaign that I finally thought of someone [_{RC} to spearhead __]

Given that maintaining a filler in working memory across structurally and/or referentially complex intervening material is known to increase processing cost (Gibson, 1998; Warren & Gibson, 2002), more recent performance-based accounts for island constraints may be able to explain why judgments are often graded. The idea that island effects reflect general cognitive capacity limitations - rather than constraints imposed by the competence grammar on hierarchical phrase structure representations - is not new (e.g. Kluender & Kutas, 1993; Hawkins, 1999), and psycholinguistic evidence in its favor has been accumulating (see Alexopoulou & Keller, 2003; Hofmeister, Jaeger, Sag, Arnon, & Snider, 2007; Kluender, 2004; Sag, Hofmeister, & Snider, 2007, among others).¹

Following Kluender (2004), the perceived ungrammaticality (and uninterpretability) of sentences containing RC islands such as (7), for example, results from processing overload at the RC clause boundary.

- (7) *We liked the book that the author [_{RC} who wrote ____] gave an interview yesterday.

In (7) above, the filler *the book* must be stored in working memory until a suitable gap has been identified in the subsequent input. At the second RC's clause boundary, the cost of maintaining the filler rises due to the parser's attempting to identify and access a discourse referent for the definite noun phrase *the author* (which at this point still lacks a thematic role and thus cannot be fully integrated into the emerging sentence representation). In addition, a semantic link must be formed between *the author* and the relative pronoun *who*, which itself

¹ Chomsky (2007: 15f.) also concedes that island constraints “may reduce in large measure to minimal search conditions of optimal computation, perhaps not coded in UG but more general laws of nature...”.

is another *wh*-filler triggering a new gap search, besides indicating the start of the new sub-clause. As a result of the increased referential processing load and memory burden, the original gap search is likely to be abandoned or suspended at this point, preventing the filler *the book* from being associated with the embedded verb *wrote*.²

In view of the above controversy surrounding the nature of island constraints, investigating non-native speakers' sensitivity to islands in L2 processing will not, by itself, be able to provide evidence for or against the SSH. Results indicating that L2 learners respect extraction islands during processing would be compatible both with processing-based and grammar-based accounts for islands, and consistent with the SSH from the point of view of the former, whereas insensitivity to islands in L2 processing would be compatible with the SSH under grammar-based approaches to islands.

Examining the presence and timing of island effects in non-native processing might, however, yield data that bears on the issue of grammar versus processing-based accounts for island constraints. Demonstrating online sensitivity to islands in L2 comprehenders would, of course, be compatible with both grammar-based and processing-based accounts for islands. From a grammar-based perspective, one would interpret island sensitivity as evidence for native-like grammatical competence in this domain. From a processing perspective, *island sensitivity* in L2 comprehension would be interpreted as resulting from the (potentially elevated) cognitive resource demands associated with processing a non-native language. That is, given that non-native speakers are known to suffer from working memory limitations in their L2 in comparison to their L1 (Miyake & Friedman, 1998; Service, Simola, Metsänheimo, & Maury, 2002), processing-based accounts for islands lead us to expect that learners should show immediate sensitivity to islands during L2 sentence comprehension, irrespective of whether or not overt *wh*-movement is instantiated in their L1.

The alternative finding of learners being *insensitive* to islands in L2 processing, however, would be more in line with grammar-based than with processing-based accounts. From a processing perspective, insensitivity to islands in non-native comprehenders would be unexpected because a failure to respect islands carries the risk of processing overload and consequent processing breakdown. Grammar-based accounts for islands, in contrast, give rise to the alternative predictions that learners might violate island constraints in their L2, or -

² Increased referential processing load and the need for long-term memory storage are properties that sentences containing extraction islands share with center-embeddings, which are well known to cause processing difficulty and elicit graded acceptability ratings (for some discussion, see Sag et al., 2007).

given earlier findings indicating that L2 learners have difficulty building complex structural representations of the input in real time - that island effects might be delayed in L2 compared to L1 processing. Grammar-based would furthermore be consistent with L1 effects on L2 learners' sensitivity to islands. Suppose, for example, that learners from *wh*-in-situ (but not those from *wh*-movement) backgrounds have non-native-like grammatical representations of *wh*-dependencies (e.g. Hawkins & Hattori, 2006). If island constraints are grammatical in nature, then learners whose L1 lacks English-type *wh*-movement should be insensitive to islands during L2 processing, whereas those from *wh*-movement backgrounds should show native-like sensitivity to islands.

2 Island constraints in processing

The results from several monolingual processing studies indicate that the L1 parser is sensitive to island constraints in that it does not attempt to link a *wh*-filler to a potential gap inside an island (see, among others, Stowe, 1986; Traxler & Pickering, 1996). However, previous findings have been somewhat mixed, with the results from some studies suggesting that certain types of island constraint may be violated in native language processing (see Phillips, 2006, for review and discussion).

The current study is based on an earlier eye-movement monitoring study by Traxler and Pickering (1996). Eye-movement records provide a rich source of information about the time-course of language processing at any given point in a sentence. Using a plausibility manipulation as a diagnostic for dependency formation, Traxler and Pickering investigated the timing of island effects in sentences containing RC islands such as (8b) below.

- (8) a. We like the *book/city* that the author wrote (____ #1) unceasingly and with great dedication about ____ #2 while waiting for a contract.
- b. We like the *book/city* that the author who wrote (____ #1) unceasingly and with great dedication saw ____ #2 while waiting for a contract.

In the absence of any extraction islands in (8a), the first potential (albeit ultimately incorrect!) gap is at the direct object position of *wrote* (= #1). Since the filler *the book* is a plausible object of *wrote* whereas *the city* is not, the parser's attempt to link the filler to *wrote* should result in elevated reading times in the 'implausible' (*the city*) compared to the 'plausible' (*the book*) condition at this point during processing. If initial *wh*-dependency formation is not

restricted by island constraints, then the same kind of plausibility effects should also be seen in (8b). If, however, the gap search is temporarily suspended during the processing of the RC island in (8b), the plausibility manipulation should not affect participants' reading times at or around *wrote* in sentence of the latter type.

Traxler and Pickering found plausibility effects only for non-island sentences such as (8a), suggesting that the L1 parser respects RC islands during processing. In addition, the authors found a reverse plausibility effect at the ultimate gap site (= #2) for sentences of type (8a), which contain a locally legitimate but globally inappropriate pseudo-gap. The reverse plausibility effect indicates that reanalysing an initially plausible dependency (i.e., *wrote the book*) is more difficult than reanalysing an initially implausible one (i.e., *wrote the city*). Together, Traxler and Pickering's findings provide strong evidence for immediate sensitivity to RC islands in native sentence processing.

Evidence from previous L2 processing studies suggests that learners, too, are sensitive to constraints on *wh*-movement in processing tasks (Juffs, 2005; Juffs & Harrington, 1995; Omaki & Schulz, 2008). Juffs and Harrington, for instance, showed that learners from different L1 backgrounds are sensitive to subjacency violations in online grammaticality judgment, while they seemed to have difficulty processing certain types of grammatical *wh*-structures that involve reanalysis of gaps. In a self-paced reading study using materials very similar to those of Traxler and Pickering, Omaki and Schulz report that proficient Spanish-speaking learners of English behaved like the native English-speaking controls in that they did not attempt to link a filler to a potential gap inside an RC island. The relative *timing* of island effects in L2 compared to L1 sentence comprehension, or the possible role of L1 background in island sensitivity, have not previously been systematically investigated, however. Building on and extending previous research on the L2 processing of filler-gap dependencies, the present study seeks to address the following specific research questions:

- Are there any differences in the timing of island effects between native and non-native English speakers?
- Does the complexity of the linguistic material between filler and gap (presence vs. absence of an intervening *wh*-clause) affect how *wh*-dependencies are resolved during native and non-native language processing?
- Do the syntactic properties of the first language affect how sensitive non-native speakers from different language backgrounds are to island constraints during the processing of *wh*-dependencies in L2 English?

3 The current study

To examine native and non-native English speakers' sensitivity to RC islands during online sentence processing we monitored participants eye-movements during reading, a time-course sensitive experimental technique that has proved suitable for the study of non-native language processing (Felser, Sato, & Bertenshaw, 2009; Frenck-Mestre, 2005; Roberts, Gullberg, & Indefrey, 2008). We created a series of short texts similar to those used in Traxler and Pickering's (1996) study. Two factors, plausibility and the presence vs. absence of *wh*-islands, were manipulated in a 2x2 design yielding four experimental conditions. An example set of texts is shown (9).

- (9) The big city was a fascinating topic for the new book.
- a. *Non-Island Constraint, Plausible*
Everyone liked the book that the author wrote continuously and with exceptionally great skill about whilst waiting for a contract.
 - b. *Non-Island Constraint, Implausible*
Everyone liked the city that the author wrote continuously and with exceptionally great skill about whilst waiting for a contract.
 - c. *Island Constraint, Plausible*
Everyone liked the book that the author who wrote continuously and with exceptionally great skill saw whilst waiting for a contract.
 - d. *Island Constraint, Implausible*
Everyone liked the city that the author who wrote continuously and with exceptionally great skill saw whilst waiting for a contract.

Each text consisted of a lead-in and a critical sentence. Plausibility was manipulated in the critical sentences and used as a diagnostic for dependency formation at two potential gap sites as in Traxler and Pickering (1996). The first potential gap site is at the verb *wrote* in (9), where dependency formation should be blocked by the presence of the *wh*-operator *who* in the two island conditions (9c,d) but not in the non-island conditions (9a,b). At the ultimately correct gap site, i.e. at the preposition *about* in (9a,b) and at the verb *saw* in (9c,d), all sentences are globally plausible. The lead-in sentences were included to increase readability, and always mentioned both the plausible and implausible manipulated NPs. Across all items,

the ordering in which these two NPs occurred was counter-balanced, such that half contained the implausible NP followed by the plausible NP, and half contained the plausible NP followed by the implausible NP.

We expect the native speakers to replicate Traxler and Pickering's (1996) finding that dependency formation occurs in non-island environments only. That is, longer reading times are expected for implausible than plausible sentences at the first potential gap site (around *wrote* in (9)) in the non-island conditions only. At the second, ultimate gap site (around *about/saw* in (9)) a reversal of this initial plausibility effect should be observed, indicating online reanalysis, with longer reading times for initially plausible than for initially implausible sentences in the non-island conditions only. No differences between the two island constraint conditions are expected at either point in the sentence.

With regard to non-native speakers' sensitivity to island constraints, we make the following predictions.

- If learners are sensitive to RC islands during L2 processing, they should pattern with the native speakers in showing plausibility effects at the first potential gap site in non-island environments only, with the possibility that these effects may be delayed in non-native processing.
- If learners are not sensitive to RC islands during processing, they should show main effects of plausibility that are not modulated by the presence or absence of islands.
- If the presence or absence of overt *wh*-movement in the L1 affect L2 processing, the German learners should pattern with the native English speakers, while the Chinese might not show any sensitivity to extraction islands.
- If learners have difficulty with online reanalysis, then the presence of a locally legitimate pseudo-gap in the non-island sentences might lead to reanalysis difficulty at the ultimate gap site, reflected in the absence of the reversed plausibility effect observed by Traxler and Pickering (1996) at the second critical region (*about whilst*).
- If the structural complexity of the material intervening between the filler and ultimate gap site affects non-native speakers' ability to link the filler to its lexical licenser, they might show longer reading times in the island compared to the non-island conditions at the second critical region (*about/saw whilst*).

Method

Participants

39 native English speakers (11 males, mean age 23.7), 26 Chinese learners (5 males, mean age 27.5) and 24 German learners (7 males, mean age 22.6) of L2 English were paid to participate in the experiment. All participants had normal or corrected to normal vision and were recruited from the University of Essex community. A summary of the learners' English proficiency scores, as gauged by the Quick Placement Test (Oxford University Press, 2001), and their bio-data is shown in Table 1. All learners scored in the 'upper intermediate' proficiency bracket or above in the Quick Placement Test.

Table 1: Summary of L2 learners' bio-data and proficiency scores

	Chinese Learners		German Learners	
	Mean	<i>SD</i>	Mean	<i>SD</i>
Age of first exposure to English	12.1	1.2	11.1	1.6
Length of immersion (years)	2.1	3.0	3.9	4.6
Quick OPT Score (%)	71.0	10.0	85.0	11.5

Materials

28 sets of experimental items were constructed as illustrated in (9).³ The manipulated NPs in the plausible and implausible conditions were matched for length and word form frequency using the CELEX lexical database (Baayen, Piepenbrock, & van Rijn, 1993), and the lexical material at the second gap site (*saw/about*) was matched between the island and non-island conditions for length. The subcategorisation biases of the verbs at the first potential gap site were assessed using a sentence completion task (compare Trueswell, Tanenhaus, & Kello, 1993). 16 native English speakers were provided with a list of sentence fragments consisting of a proper noun followed by a potentially transitive verb (e.g. *John hunted*) and were asked

³ A full list of experimental items can be made available on request by C. Felser (*felser@essex.ac.uk*).

to complete the fragment with the first appropriate continuation that came to mind. To increase the likelihood of a (temporary) direct object reading of the manipulated NP at the first potential gap site, only verbs which had the highest proportion of direct object continuations (mean 69%, range 50-92%) were used to construct the experimental materials.

The strength of the plausibility manipulation was also pre-tested. 10 native English speakers rated 72 sentences (e.g. ‘the author wrote the book/city’) on a scale from 1 (plausible) to 5 (implausible). Two presentation lists were constructed, so that participants saw each sentence only once, in either the plausible or implausible version. Items were then selected on the basis of these scores, rejecting those with intermediate ratings. Mean ratings for the 28 selected items were 4.49 for the implausible NPs and 1.21 for the plausible NPs. This difference was highly significant ($t_1(9) = 17.05, p < .001$; $t_2(27) = 29.69, p < .001$). 10 (different) native English speakers also took part in an additional pre-test to ensure that all the experimental items were equally plausible globally at the second gap site. They rated sentences such as ‘the author wrote about the book/city’. Again, two item lists including 72 sentences were constructed. Plausibility ratings for the final set of 28 items did not differ between implausible (mean rating 2.1) and plausible (mean rating 2.2) conditions ($t_1(9) < 1, p = 0.613$; $t_2(27) < 1, p = 0.475$).

In addition to the 28 experimental items, 32 fillers were also constructed. 10 of these were structurally similar pseudo-fillers, five of which were globally plausible, and five mildly implausible. Of the remaining 24 fillers, a further five were mildly implausible globally. Globally implausible fillers were included in order to discourage participants from adopting a strategy of ignoring the critical sentences that were occasionally initially implausible, but always ultimately plausible globally.

Procedure

The order of the experimental and filler items was pseudo-randomized such that no two items from the same condition appeared adjacent to each other, in four presentation lists divided across participants in a Latin-square design. The experiment began with five practice items to familiarize participants with the procedure. All items were presented in Courier New font, and displayed across three lines of text on a computer screen.

Eye-movements were recorded using the head-mounted EYELINK II system. The eye-tracker records participants’ eye-movements via two cameras mounted on a headband at a sample rate of 500Hz, with spatial accuracy better than 0.5°. Participant head movements are automatically compensated for by a third camera mounted in the centre of the headband,

which tracks the position of four LEDs mounted on the corner of the computer screen. While viewing was binocular, the eye-movement data were recorded from the right eye only.

An experimental session began with the calibration of the eye-tracker on a nine-point grid. Prior to the presentation of each trial, calibration was checked via presentation of a drift correct marker above the first word of the next trial to be displayed. Participants were instructed to fixate upon this marker, and press a button to view the next trial. Any drift in the headset was automatically compensated for before presentation of the trial.

Participants were instructed to read the sentences silently at their normal reading rate, and press a button once completed. To ensure that participants paid attention to the content of the sentences, two-thirds of all trials were followed by a yes-no comprehension question, half of which required a 'yes' response, and half a 'no' response. The native English speakers completed the experiment in one session lasting approximately 30-40 minutes, while the L2 learners required two sessions of similar length. In the first session the L2 learners completed the main eye-movement experiment followed by a vocabulary checklist, and in the second they completed the proficiency test. The vocabulary checklist contained critical vocabulary from the experimental items, and in particular included the manipulated NPs and critical verbs. The L2 learners simply had to go through the list and circle any words that they were unfamiliar with.

Data analysis

To examine dependency formation at the two potential gap sites, we report reading times for two regions of text. The *first critical region* consisted of the verb at the first potential gap site plus the following word (*wrote continuously* in (9) above), while the *second critical region* was the verb or preposition at the second gap site, again plus the following word (*about whilst* and *saw whilst* in (9)). Four reading time measures were calculated at these two regions. *First-pass reading time* is the summed duration of all fixations within a region during its first inspection, including all fixations starting when the eyes first enter the region from the left, up until it is exited to either the left or right. *Regression path duration* is the sum of all fixations, starting with the first fixation when a region is first entered from the left, up until but not including the first fixation when that region is first exited to the right. Collectively, these two measures will be referred to as *first-pass* measures, and can be contrasted to the following measures that include (additional) *second-pass* processing. *Rereading time* is the summed duration of all fixations within a region *after* it has been exited to either the left or right for the first time, while *total viewing time* is the summed duration of

all fixations within a region. Reading times for trials in which track loss occurred or in which a region was initially skipped were treated as missing data. For rereading time, trials in which a region was not refixated following the first-pass contributed a rereading of zero to the calculation of averages.

Prior to the calculation of these measures an automatic procedure merged short fixations of 80ms or below that were within one degree of visual arc of another fixation. All other fixations of 80ms or below, as well as those above 800ms, were removed before any further analysis. Outliers above or below 2.5SDs of the participant mean for each measure at each region were also removed before the statistical analysis.

Results

Overall accuracy to the comprehension questions was 86%, 85% and 84% for the English native speakers, the Chinese learners and the German learners respectively, indicating that participants paid attention to the content of the sentences. Track loss accounted for 0.7% of the English data, 5.08% of the Chinese data and 0.89% of the German data. An additional 3.06% of the Chinese data and 1.94% of the German data were also removed following vocabulary screening, as participants indicated not knowing the manipulated NPs or critical verbs. Skipping rates for the two reported regions in all groups were below 3.6%, and outlier removal resulted in the loss of no more than 4.3% of the data per measure and region for each group.

A series of 2x2x3 ANOVAs with Plausibility, Island Constraint and Language Group as independent variables were undertaken for each reading time measure at both regions of text. Significant main effects of Language Group (all $ps < .001$) were found for all measures at both regions of text, indicating that the learners generally read the sentences more slowly than the native English speakers, and will not be discussed further below.

The reading time data for the *first critical region* are presented in Table 2. Analysis of the first-pass reading times for this region revealed a significant main effect of Island Constraint ($F_1(1, 86) = 24.89, p < .001$; $F_2(1, 27) = 15.35, p = .001$), indicating that participants spent more time reading non-island constraint than island constraint sentences during the initial inspection of this region, and a main effect of Plausibility by items ($F_1(1, 86) = 2.00, p = .161$; $F_2(1, 27) = 6.34, p = .018$), but no significant interactions. There was also a reliable main effect of Island Constraint by items in the regression path times ($F_1(1, 86) = 1.85, p = .177$; $F_2(1, 27) = 4.47, p = .044$), as well as an Island Constraint by Language Group interaction, again only by items ($F_1(2, 86) = 1.03, p = .353$; $F_2(2, 54) = 3.24, p =$

.047). As in the first-pass times, participants tended to spend longer in the non-island constraint than the island constraint sentences, with the effect in this measure being mainly carried by the German group. There were no other reliable main effects or interactions.

The analysis of the rereading times revealed a significant main effect of Plausibility ($F_1(1, 86) = 7.47, p = .008$; $F_2(1, 27) = 9.32, p = .005$), as well as an Island Constraint by Plausibility interaction in the analysis by participants that was marginal by items ($F_1(2, 86) = 10.50, p = .002$; $F_2(1, 27) = 3.74, p = .064$), but no other reliable main effects or interactions. To examine the Island Constraint by Plausibility interaction, in the absence of any interactions with language group, t-tests were carried out on the lumped means of the participants as one large group as a whole. These revealed no reliable differences between the two island conditions ($t_1(88) < 1, p = .966$; $t_2(27) < 1, p < .762$), but the implausible non-island condition had longer rereading times than the plausible non-island condition ($t_1(88) = 3.47, p = .001$; $t_2(27) = 3.67, p = .001$).

Much the same pattern of results was found in the total viewing times. There was a significant main effect of Plausibility ($F_1(1, 86) = 14.7, p < .001$; $F_2(1, 27) = 14.89, p = .001$), and an Island Constraint by Plausibility interaction ($F_1(2, 86) = 5.64, p = .020$; $F_2(1, 27) = 3.98, p = .056$) in the absence of any other reliable main effects or interactions. Lumped t-tests again revealed no reliable differences between the two island conditions ($t_1(88) = 1.37, p = .174$; $t_2(27) = 1.03, p = .315$), but the implausible non-island condition had longer total viewing times than the plausible non-island condition ($t_1(88) = 3.68, p < .001$; $t_2(27) = 4.23, p < .001$).

The results for the first critical region can be summarised as follows. Readers initially spent more time reading non-island than island constraint sentences. Although no reliable plausibility effects were found in either of the two first-pass measures, they were found in both rereading and total viewing times. In these two measures, the participants as a whole exhibited longer reading times for implausible than plausible sentences, in non-island constraint conditions only. We found no reliable evidence of plausibility effects for any group in the island constraint conditions. This indicates that none of our participant groups attempted to form dependencies in island environments, suggesting that the formation of filler-gap dependencies in both native and non-native English is restricted by island constraints during online processing.

Table 2: Mean durations in msec (and standard deviations) of four reading time measures at the first critical region

	First-Pass		Regression Path		Rereading		Total Viewing	
	Reading Time		Duration		Time		Time	
<i>Native English Speakers</i>								
Non-Island Constraint/Plausible	564	(144)	689	(191)	263	(238)	834	(267)
Non-Island Constraint/Implausible	533	(134)	722	(206)	412	(375)	964	(416)
Island Constraint/Plausible	471	(157)	703	(221)	339	(307)	825	(355)
Island Constraint/Implausible	484	(136)	655	(187)	344	(235)	840	(283)
<i>Chinese Learners</i>								
Non-Island Constraint/Plausible	852	(272)	1174	(430)	791	(629)	1674	(780)
Non-Island Constraint/Implausible	904	(261)	1230	(436)	1094	(907)	1992	(906)
Island Constraint/Plausible	802	(144)	1165	(452)	950	(623)	1761	(626)
Island Constraint/Implausible	831	(269)	1257	(570)	960	(648)	1833	(716)
<i>German Learners</i>								
Non-Island Constraint/Plausible	595	(208)	780	(270)	452	(408)	1067	(487)
Non-Island Constraint/Implausible	683	(239)	819	(279)	467	(437)	1179	(532)
Island Constraint/Plausible	573	(167)	706	(265)	501	(483)	1084	(521)
Island Constraint/Implausible	554	(173)	705	(230)	487	(400)	1146	(528)

Table 3: Mean durations in msecs (and standard deviations) of four reading time measures at the second critical region

	First-Pass		Regression Path		Rereading		Total Viewing	
	Reading Time		Duration		Time		Time	
<i>Native English Speakers</i>								
Non-Island Constraint/Plausible	444	(128)	577	(212)	271	(228)	737	(297)
Non-Island Constraint/Implausible	472	(149)	550	(195)	207	(163)	698	(266)
Island Constraint/Plausible	480	(160)	576	(193)	280	(255)	773	(329)
Island Constraint/Implausible	485	(140)	590	(190)	264	(190)	766	(271)
<i>Chinese Learners</i>								
Non-Island Constraint/Plausible	647	(227)	834	(332)	601	(436)	1277	(584)
Non-Island Constraint/Implausible	613	(168)	760	(199)	544	(393)	1181	(421)
Island Constraint/Plausible	754	(207)	1061	(534)	835	(681)	1581	(711)
Island Constraint/Implausible	757	(235)	1079	(508)	738	(579)	1558	(620)
<i>German Learners</i>								
Non-Island Constraint/Plausible	536	(192)	711	(317)	438	(411)	989	(489)
Non-Island Constraint/Implausible	557	(218)	658	(315)	363	(382)	930	(494)
Island Constraint/Plausible	598	(183)	725	(292)	539	(541)	1159	(586)
Island Constraint/Implausible	588	(170)	779	(294)	559	(569)	1147	(666)

The results of the *second critical region* can be found in Table 3. Analysis of this region's first-pass reading times revealed a significant main effect of Island Constraint ($F_1(1, 86) = 31.36, p < .001$; $F_2(1, 27) = 12.42, p = .002$), and an Island Constraint by Language Group interaction by participants ($F_1(2, 86) = 7.11, p = .001$; $F_2(2, 54) = 2.37, p = .104$), but no other reliable main effects or interactions. Participants generally had longer first-pass reading times for the second critical region in the island than the non-island constraint conditions. The Island Constraint by Language Group interaction indicates that the size of this effect was larger for the non-native groups, and in particular the Chinese learners, than the native English speakers.

The regression path times also showed a reliable main effect of Island Constraint ($F_1(1, 86) = 23.47, p < .001$; $F_2(1, 27) = 16.44, p < .001$), an Island Constraint by Language Group interaction ($F_1(2, 86) = 9.71 < .001, p = .001$; $F_2(2, 54) = 6.53, p = .003$), and an Island Constraint by Plausibility interaction ($F_1(2, 86) = 6.05, p = .016$; $F_2(1, 27) = 8.90, p = .006$). No other main effects or interactions were reliable. To examine the Island Constraint by Plausibility interaction, in the presence of an interaction with Language Group, t-tests were carried out on each language group separately. These however, revealed no reliable differences between either the two island conditions, nor the two non-island conditions for any of the three language groups.

The rereading times again showed a main effect of Island Constraint ($F_1(1, 86) = 25.20, p < .001$; $F_2(1, 27) = 21.99, p = .002$), and an Island Constraint by Language Group interaction that was marginal by items ($F_1(2, 86) = 4.62, p = .012$; $F_2(2, 54) = 2.66, p = .079$). There was also a significant main effect of Plausibility by participants ($F_1(1, 86) = 5.98, p = .016$; $F_2(1, 27) = 2.89, p = .101$), but no other reliable interactions. To examine the plausibility effect in the presence of an interaction with Language Group, t-tests were again carried out on each group separately. The native English speakers showed no reliable differences between the two island conditions ($t_1(38) < 1, p = .610$; $t_2(27) < 1, p = .681$), but the rereading times for the plausible non-island condition were reliably longer than those for the implausible non-island condition by participants, and marginally so by items ($t_1(38) = 2.62, p = .012$; $t_2(27) = 1.84, p = .077$). Neither learner group showed any reliable differences in these comparisons.

Finally, the total viewing times also showed a significant main effect of Island Constraint ($F_1(1, 86) = 55.52, p < .001$; $F_2(1, 27) = 35.97, p = .002$), a reliable Island Constraint by Language Group interaction ($F_1(2, 86) = 11.21, p < .001$; $F_2(2, 54) = 5.60, p = .006$), a marginally significant main effect of Plausibility by participants ($F_1(1, 86) = 3.47, p = .066$; $F_2(1, 27) = 2.09, p = .159$), but no other reliable interactions. T-tests were carried out to explore the plausibility effect in each language group separately, but these revealed no reliable differences between either the two island conditions, nor the two non-island conditions for any group.

The reading time data at the second critical region, where the filler-gap dependency is ultimately resolved, showed reliable Island Constraint by Language Group interactions in every reported measure. The L2 learners had longer reading times for the island than the non-island conditions at this point in the sentence, in comparison to the native English speakers. This finding would suggest that the presence of an intervening RC island makes it particularly difficult for non-native speakers to link a distant *wh*-filler to its lexical subcategorizer during online processing. Additionally, the rereading times of the native English speakers were found to be longer for initially plausible than initially implausible sentences, in non-island environments only. This finding provides evidence of online reanalysis of the initial, incorrectly posited dependency that was formed at the first critical region, when the native speakers encountered the ultimate gap site. No reliable evidence of online reanalysis of the initial (incorrect) dependency formed at the first critical region was observed for either group of L2 learners at the ultimate gap site, however.

4 Discussion

The results from the native English speakers replicated Traxler and Pickering's (1996) earlier findings. Our plausibility diagnostic showed that participants temporarily postulated a *wh*-gap in non-island environments only, and that a *wh*-dependency that was plausible at the first potential gap site was more difficult to revise at the ultimate gap site than one that was implausible to start with. The L2 learners patterned with the native speakers in showing evidence for dependency formation only in non-island environments, and there were no statistical differences in the timing of island constraints in native

compared to non-native sentence processing. If anything, the predicted reading-time pattern was seen earlier in the learners than in the native controls numerically, especially in the German group, who showed an 88ms advantage for plausible compared to implausible *wh*-fillers in their first-pass reading times at the first critical region in the non-island conditions. This is consistent with earlier findings from self-paced reading and online plausibility judgment tasks indicating that non-native speakers are able to link a filler to a potential lexical subcategorizer during online processing (Marinis et al., 2005; Williams et al., 2001; Williams, 2006).

Unlike the native English controls, however, the learners showed no reliable evidence for online reanalysis at the ultimate gap site. This is in line with previous observations suggesting that learners may have more difficulty than native speakers revising erroneous initial parsing decisions online (Hopp, 2006; Juffs & Harrington, 1995 – but see Williams, 2006). Instead, the two L2 groups, in particular the Chinese learners, had longer reading times for island sentences than for non-island sentences generally. This finding suggest that the L2 learners had difficulty ultimately resolving the original *wh*-dependency in the presence of an intervening RC clause, which renders the experimental sentences in the island conditions structurally and pragmatic more complex than our non-island sentences. This is unsurprising both from the point of view of the SSH, which claims that non-native speakers have difficulty building complex structural representations in real time (compare e.g. Marinis et al., 2005), and considering the high referential processing and memory load associated with sentences containing extraction islands. That is, the unusually high processing effort required at the RC clause boundary makes it difficult, especially for non-native comprehenders, to maintain the original *wh*-filler in memory and/or retrieve it from memory at its ultimate gap site.

Our results furthermore show that typological L1/L2 distance - specifically, the presence versus absence of English-type *wh*-movement in the learners' first language - did not affect their sensitivity to extraction islands in L2 processing. Learners from both *wh*-movement (German) and *wh*-in-situ (Chinese) backgrounds applied island constraints online to restrict the formation of filler-gap dependencies. In view of evidence suggesting that learners from *wh*-in-situ backgrounds may have representational deficits in the

domain of *wh*-dependencies (e.g. Hawkins & Hattori, 2006), this finding would also seem to be more compatible with a processing-based account for islands.

As pointed out earlier, the results from the current study cannot by themselves provide any direct evidence for or against the SSH, given the current debate as to whether island effects reflect processing capacity limitations or grammatical constraints on hierarchical phrase structure configurations. Our results clearly are compatible with the SSH under a processing-based account. Alternatively, however, under the assumption that island constraints are grammatical rather than processing-based, learners' immediate sensitivity to islands in online processing reflects native-like grammatical competence and parsing abilities (contra the SSH). Note that this assumption fails to account, however, not only for the learners' difficulty establishing *wh*-dependencies across an intervening subordinate clause observed in the current study, but also for the apparent absence of intermediate gaps in learners' online representations in Marinis et al.'s (2005) and Pliatsikas and Marinis's (2009) studies.

Assuming a performance-based account for islands, on the other hand, provides us with a rather more coherent 'big picture'. Under this view, our finding that learners from typologically different L1 backgrounds whose general English grammar proficiency was good but by no means native-like showed immediate sensitivity to RC islands is entirely unsurprising. If dependencies between a filler and a verb inside a relative clause are not formed as a result of referential processing and memory overload at clause boundaries, non-native speakers will be as unlikely as native speakers to attempt to form dependencies inside island clauses, regardless of their L1 background, and even at intermediate levels of L2 grammar proficiency. The postulation of intermediate structural gaps at the left clausal periphery, in contrast, requires comprehenders to assemble fully fledged hierarchical phrase structure representations. If learners have difficulty building complex structural representations in real time, as claimed by the SSH, then the structural scaffolding required to host an intermediate gap is likely to be missing from learners' online sentence representations, as Marinis and colleagues' results indicate.

5 Conclusion

Taken together, our results suggest that both native and non-native readers are sensitive to RC islands during language comprehension, but that non-native speakers have more difficulty than native speakers linking distant *wh*-fillers to their lexical subcategorizers in the presence of an intervening RC island. Our findings lend support to the claim that island effects are performance-based and attributable to the increased processing load at RC clause boundaries. By increasing global sentence complexity, the presence of an intervening island clause makes it difficult especially for non-native comprehenders to maintain the filler in memory and/or retrieve it again at the point of the ultimate gap.

Acknowledgments

This research was supported by an ESRC grant (RES-000-22-2508) to C. Felser. We thank the audiences at the *GASLA 10* (University of Illinois at Urbana-Champaign, March 2009), *Formal vs. Processing Explanations of Syntactic Phenomena* (University of York, April 2009) and *Second Language Processing and Parsing* (Texas Tech University, May 2009) conferences for helpful comments and discussion.

References

- Alexopoulou, D., & Keller, F. (2003). Linguistic complexity, locality and resumption. In L. Mikkelsen, & C. Potts (Eds.), *Proceedings of WCCFL 21* (pp.15-28). Somerville, MA: Cascadilla Press.
- Baayen, H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX lexical database (CD ROM). Philadelphia, PE: Linguistic Data Consortium, University of Pennsylvania.
- Chomsky, N. (1973). Conditions on transformations. In S. Anderson, & P. Kiparsky (Eds.), *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston.
- Chomsky, N. (2007). Of minds and language. *Biolinguistics*, 1, 9-27.
- Chung, S., & McCloskey, J. (1983). On the interpretation of certain island facts in GPSG. *Linguistic Inquiry*, 14, 704-713.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3-42.
- 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., Sato, M., & Bertenshaw, N. (2009). The on-line application of Binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, 12.
- Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93-126.
- Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research*, 21, 175-198.
- Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-75.
- Hawkins, J. (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., T.F. Jaeger, I. Sag, I. Arnon, & N. Snider (2007). Locality and accessibility in *wh*-questions. In S. Featherston, & W. Sternefeld (Eds.), *Roots: Linguistics in Search of its Evidential Base* (pp. 185-206). Berlin: Mouton de Gruyter.
- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22, 369-397.
- 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.
- Kluender, R. (2004). Are subject islands subject to a processing account? In A. Rodriguez, V. Chand, A. Kelleher, & B. Scheiser (Eds.), *Proceedings of WCCFL* 23 (pp. 475-499). Somerville, MA: Cascadilla Press.
- Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8, 573-633.
- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27, 53-78.
- Miyake, A., & Friedman, N. (1998). Individual differences in second language proficiency: working memory as language aptitude. In A. Healy, & L. Bourne (Eds.), *Foreign language learning: psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum Associates.
- Omaki, A., & Schulz, B. (2008). Evidence for the online use of island constraints in second language processing. Paper presented at Second Language Research Forum 2008, Honolulu, 17 October.
- Phillips, C. (2006). The real-time status of island phenomena. *Language*, 82, 795-823.
- Pliatsikas, C., & Marinis, T. (2009). Does type of exposure matter in processing long-distance *wh*-movement by late L2 learners? Paper presented at the L2 Processing and Parsing Conference, Texas Tech University, 21-24 May 2009.
- Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, 30, 333-357.
- Ross, J.R. (1967). *Constraints on variables in syntax*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
- Sag, I., Hofmeister, P., & Snider, N. (2007). Processing complexity in subjacency violations: the complex noun phrase constraint. To appear in *Proceedings of 43rd Regional Meeting of the Chicago Linguistics Society*.
- Service, E., Simola, M., Metsänheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, 14, 383-408.
- Staub, A., & Rayner, K. (2007). Eye movements and online comprehension processes. In M. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 327-342). Oxford: Oxford University Press.
- Stowe, L. (1986). Parsing *wh*-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227-245.
- Traxler, M., & Pickering, M. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35, 542-562.

- Trueswell, J., Tanenhaus, M., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 528-553.
- Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, 85, 79-112.
- Williams, J., Möbius, P., & Kim, C. (2001). Native and non-native processing of English *wh*-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, 22, 509-540.
- Williams, J. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, 9, 71-88.