Bilingualism: Language and Cognition 12 (4), 2009, 485–502 © Cambridge University Press 2009 doi:10.1017/S1366728909990228
First published online 24 August 2009

The on-line application of binding Principle A in English as a second language*

CLAUDIA FELSER
MIKAKO SATO
NICHOLAS BERTENSHAW
University of Essex

We report the results from two experiments investigating proficient Japanese-speaking learners' processing of reflexive object pronouns in English as a second language (L2). Experiment 1 used a timed grammaticality judgement task to assess learners' sensitivity to binding Principle A under processing pressure, and Experiment 2 investigated the time-course of reflexive anaphor resolution during L2 reading using the eye-movement monitoring technique. Taken together, our results show that despite having demonstrated native-like knowledge of reflexive binding in corresponding untimed tasks, the learners processed English reflexives differently from native speakers in that they took into consideration a matching discourse-prominent but binding-theoretically inappropriate antecedent when first encountering a reflexive. This suggests that unlike what has been reported in corresponding monolingual processing research (Sturt, 2003), initial antecedent search in L2 English is not constrained by binding Principle A.

1. Introduction

The ability to link anaphoric elements such as reflexive pronouns to their antecedents during listening or reading is a crucial prerequisite for successful sentence and discourse comprehension. Given that this ability is just as important for second language (L2) users as it is for native speakers, the current lack of studies investigating the L2 processing of anaphoric expressions is surprising. Several L2 processing studies have shown that even learners who are indistinguishable from native speakers in off-line tasks may perform differently from the latter in on-line processing tasks, and there is evidence that discontinuous syntactic dependencies are particularly difficult for learners to establish during L2 processing (Marinis, Roberts, Felser and Clahsen, 2005; Felser and Roberts, 2007). This, in turn, has been argued to reflect learners' reduced ability to build structurally complex, native-like grammatical representations during L2 processing and a corresponding over-reliance on non-structural cues to interpretation (Clahsen and Felser, 2006). The present study aims to further explore the degree to which non-native comprehenders are able to establish grammatically mediated dependencies in L2 sentence processing by examining their sensitivity to syntactic coreference constraints on reflexives.

* The research reported here has been supported by a University of Essex Research Promotion Fund grant to C. Felser and by an ESRC postgraduate studentship to N. Bertenshaw. We also thank Harald Clahsen, the audience at the November 2007 Workshop on "Language Processing in First and Second Language Learners" at the Max Planck Institute for Psycholinguistics, Nijmegen, and four anonymous *BLC* reviewers for helpful comments and suggestions.

Address for correspondence:
Claudia Felser, Department of Language and Linguistics, University of Essex, Colchester CO4 3SQ, UK
felsec@essex.ac.uk

According to Principle A of the binding theory, reflexive anaphors must be bound within their binding domain, that is, they must be linked to a c-commanding noun phrase that matches the reflexive in gender, person and number (Chomsky, 1981).[1] There is considerable cross-linguistic variation, however, as to what may constitute a relevant binding domain in a given language. Whereas English reflexives normally require clausemate antecedents (hence the impossibility of interpreting *himself* as coreferential with the matrix subject *John* in (1a)), the commonly used Japanese reflexive *zibun* "self", for instance, also permits long-distance (LD) binding by the subject of a higher clause, as shown in (1b).

(1) a. John_i thinks that David_k hates $\text{himself}_{*i/k}$.
b. Taroo_i-wa Ziroo_k-ga $\text{zibun}_{i/k}$-o kiratteiru-to omotteiru.
Taro-TOP Ziro-NOM self-ACC hate-COMP think
"Taro thinks that Ziro hates self."

Exactly how such cross-linguistic differences in binding domains are best accounted for remains a matter of debate within theoretical linguistics (see Büring, 2005, for review and discussion), but at the descriptive level the availability of LD binding appears to be restricted to morphologically simplex reflexives such as Japanese *zibun*, Chinese *ziji* or Dutch *zich*, whereas polymorphemic reflexives such as English *-self* or Dutch *-zelf* anaphors usually require local binding (compare e.g. Reinhart

[1] The term c-command refers to a relationship between nodes in a phrase marker that is based on the hierarchical notion of dominance. Generally speaking, a constituent c-commands its sister constituents and any constituents that these dominate (compare Reinhart, 1981).

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

and Reuland, 1993).[2] Japanese resembles languages like Dutch in that both types of reflexive are available. Although *zibun* is the primary reflexive in Japanese, polymorphemic reflexives such as *zibun-zisin* "self-self" or *kare*/*kanojo-zisin* "he/she-self" are also used, which like English reflexives are normally bound locally. The latter, which is not subject-oriented and also shows a gender contrast, is in fact the most direct translation equivalent of English reflexives.

Previous second language acquisition (SLA) research on reflexive binding has virtually exclusively been informed by data from off-line tasks such as picture identification, truth-value judgements or antecedent identification. Although the results from these studies have not been entirely consistent, there is evidence from several studies suggesting that learners from LD-binding backgrounds are able to acquire the locality requirement on English reflexives (see, among others, Thomas, 1991; Eckman, 1994; Matsumura, 1994). As Hawkins (2001, p. 312) notes, however, results from off-line tasks are often difficult to interpret because it is impossible to tell to what extent learners' judgements or antecedent choices might have been influenced by non-linguistic or 'performance' factors. It is conceivable, for example, that some learners might choose the correct antecedent in sentences such as (1a) above merely because focussing on the local antecedent is the more working-memory–friendly option, rather than as a result of having successfully 'reset' the relevant grammatical parameters. It is also possible that learners whose ultimate antecedent choices are native-like nevertheless experience some non-native-like processing difficulty – due, perhaps, to interference from their first language (L1) – prior to providing their responses.

Factors such as slower processing speed or working memory limitations are likely to have a stronger impact on learners' performance in on-line processing tasks. However, White, Bruhn-Garavito, Kawasaki, Pater and Prévost's (1997) finding that learners' performance varied considerably between a story-based and a picture-based variant of their truth-value judgement task suggests that differences in task demands may also affect L2 anaphor resolution in off-line or untimed tasks. We thus agree with White et al. (1997, p. 163) that in order to gain a better understanding of learners' ability to apply language-specific binding requirements in the L2, "convergent evidence from a variety of tasks" should be sought – including on-line processing tasks. While off-line tasks can provide useful measures of learners' grammatical knowledge and of their ultimate interpretations, the use of on-line experimental techniques can help identify areas of processing difficulty that are not necessarily obvious from learners' performance in untimed tasks, and can provide a window on the mental processes underlying learners' interpretations.

Whereas most previous SLA research has investigated L2 reflexive binding from the perspective of Chomsky's (1981) principles-and-parameters framework with the aim of uncovering details of the L2 interlanguage grammar (see Hawkins, 2001; White, 2003, for reviews), experimental psycholinguistic research on anaphor resolution has typically focused on questions such as what information sources (syntax, pragmatics, etc.) influence antecedent search and at what stage during processing these become available (see Nicol and Swinney, 2003, for a review). Our study seeks to help bridge the current research gap between the two fields by using both metalinguistic judgement and on-line reading comprehension tasks to help determine whether proficient Japanese-speaking learners of English are sensitive to binding Principle A during L2 reading. Specifically, we examine whether learners who are indistinguishable from native speakers in untimed tasks also PROCESS reflexive pronouns in the same way as the latter. To our knowledge, ours is the first study to examine the time-course of L2 reflexive anaphor resolution using the eye-movement monitoring technique.

2. Processing reflexive anaphors in the L1

The role of binding constraints in on-line reflexive anaphor resolution has previously been investigated in monolingual processing studies only (see, among others, Nicol and Swinney, 1989; Harris, Wexler and Holcomb, 2000; Badecker and Straub, 2002; Sturt, 2003; Runner, Sussman and Tanenhaus, 2003, 2006; Xiang, Dillon & Phillips, 2009). Studies using time-course sensitive measures such as eye movements or event-related potentials (ERPs) have shown that native readers quickly link argument reflexives to their binding-accessible antecedents during processing (Harris et al., 2000; Sturt, 2003; Xiang et al., to appear). Of particular interest to the current study, however, is the question of whether (and if so, when) during processing syntactically inappropriate antecedents may be considered as possible referents for reflexives.

Based on results from cross-modal priming tasks, Nicol and Swinney (1989) argue that the binding principles constrain anaphor resolution from the earliest stages of processing onwards in English monolingual sentence comprehension. They report evidence that upon hearing the reflexive in sentences such as (2) below, participants mentally reactivated only the binding-theoretically appropriate antecedent (*the doctor*) but not

[2] Although *zibun* is written as two kanji characters (*zi* "self" and *bun*, whose possible meanings include "portion", "division", and "given status"), it is commonly considered a monomorphemic element or X^0 constituent (compare e.g. Cole & Sung, 1994). There is evidence suggesting that in Japanese, LD binding of *zibun* is in fact preferred over local binding (e.g. Hirakawa, 1990; Nagata, 1995).

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

any of the other noun phrase referents mentioned (i.e. *the boxer* or *the skier*).

(2) The boxer told the skier that the doctor for the team would blame himself for the recent injury.

According to the 'binding-as-initial-filter' hypothesis, syntactic coreference constraints immediately exclude any structurally inappropriate antecedents from the candidate set, while pragmatic or discourse-related information is taken into consideration only when syntactic constraints fail to unambiguously identify the referent.

However, the 'binding-as-initial-filter' hypothesis has been called into question by the results from a series of reading-time experiments reported by Badecker and Straub (2002), which indicate that a syntactically and pragmatically salient but illegitimate (or 'binding-inaccessible') antecedent such as *John* in (3) can also affect on-line reference resolution.

(3) John (Jane) thought that Bill owed himself another opportunity to solve the problem.

The authors found that reading times for the second word following the reflexive (i.e. *opportunity* in (3)) increased significantly when the matrix subject matched the reflexive in gender compared to when it did not. On the other hand, no such differences between the 'double-match' (i.e. *John . . . Bill . . . himself*) and 'single-match' conditions (i.e. *Jane . . . Bill . . . himself*) were observed for sentences in which a matching but illegitimate antecedent was less discourse-prominent and did not c-command the reflexive.

The results from 'visual-world' eye-movement experiments by Runner et al. (2003, 2006) also appear to pose a problem for Nicol and Swinney's (1989) hypothesis as they show that reflexives in picture noun phrases containing possessors (such as *Mary's picture of herself*) also allow non-local binding, contrary to what traditional binding theory would lead us to expect. As the authors point out themselves, however, the observed ambiguity of reflexives in picture noun phrases does not necessarily present a problem for the 'binding-as-initial-filter' account if reflexives in picture noun phrases are classed as logophors rather than argument reflexives. Unlike structural reflexives, which are subject to binding Principle A, the interpretation of logophors is sensitive to pragmatic and discourse factors (compare e.g. Pollard and Sag, 1992; Reinhart and Reuland, 1993).

The relative timing of syntactic and discourse-level constraints on argument reflexives in English has been examined more closely by Sturt (2003) using eye-movement monitoring during reading, an experimental technique that provides a rich source of data on moment-to-moment language processing (compare e.g. Staub and Rayner, 2007). Sturt's materials included sentences such as (4), with both the non-local (or 'binding-inaccessible') antecedent's and the reflexive's gender being manipulated so as to create a total of four experimental conditions.

(4) Jonathan (Jennifer) was pretty worried at the City Hospital. He (she) remembered that the surgeon had pricked himself (herself) with a used syringe needle. There should be an investigation soon.

The reflexive here is bound by the local c-commanding antecedent, *the surgeon*, which denotes a stereotypically 'male' profession. Thus, the use of the feminine form *herself* creates a mismatch between the reflexive's and the binding-accessible antecedent's stereotypical gender. Although the subject pronoun *he* or *she* also c-commands the reflexive, the locality requirement on English reflexives renders it inaccessible for binding. Sturt found that participants' initial fixations of the reflexive were shorter when the reflexive matched the binding-accessible antecedent's stereotypical gender than when it did not, demonstrating that the locality constraint on binding was applied immediately upon the participants' first reading of the reflexive. Effects of the binding-inaccessible antecedent's gender were not observed until later on during processing, with second-pass reading times being longer in the pre-final sentence region when the inaccessible antecedent did not match the reflexive's gender. The early accessible antecedent-mismatch (*surgeon . . . herself*) effect was replicated in a second experiment using materials in which the binding-inaccessible antecedent was linearly closer to but failed to c-command the reflexive, such as (5).

(5) The surgeon who treated Jonathan (Jennifer) had pricked himself (herself) with a used syringe needle.

No effects of the binding-inaccessible antecedent's gender (i.e. *Jonathan* vs. *Jennifer*) were observed in either early or later eye-movement measures here, however. This confirms that the participants had not attempted to link the reflexive to the linearly closest potential antecedent but had been influenced by the binding-inaccessible antecedent's relative hierarchical or discourse prominence instead.[3]

Like the findings reported by Nicol and Swinney (1989), Sturt's eye-movement results provide evidence for the early application of Principle A during reflexive anaphor resolution in monolingual processing. According to Sturt, the delayed effect of the discourse-prominent but structurally inaccessible 'competitor' antecedent that was seen in his first experiment shows that native readers may also take into consideration a highly salient but binding-inaccessible antecedent during later processing

[3] Evidence that native speakers are unaffected by a linearly closer but binding-inaccessible antecedent during early processing stages also comes from an ERP study by Xiang et al. (to appear).

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

stages, however. Note that Sturt's 'binding-as-defeasible-filter' hypothesis would also seem to be compatible with Badecker and Straub's (2002) results, which did not show any effects of the inaccessible antecedent until two words on from the reflexive.[4]

In summary, there is evidence from several monolingual processing studies suggesting that binding constraints are operative during on-line processing, although no ultimate consensus has yet been reached as to their relative strength and timing. The current study aims to contribute to and extend this line of research by comparing whether, and at what stage during processing, native and non-native anaphor resolution is influenced by a binding-inaccessible 'competitor' antecedent.

3. Experiment 1: grammaticality judgements

Our first experiment uses both untimed and timed ('speeded') judgement tasks to examine the effect of processing pressure on learners' ability to discriminate between grammatical and ungrammatical sentences containing reflexive anaphors in L2 English. The main empirical questions to be addressed here are the following:

- Is learners' ability to judge sentences containing reflexive anaphors influenced by the binding-inaccessible antecedent's relative discourse-prominence (i.e. by whether or not it c-commands the reflexive)?
- How and to what extent do individual differences in reading span or processing speed influence participants' judgements?

Unlike untimed judgement tasks, speeded grammaticality judgement tasks with controlled stimulus presentation give participants little time to reflect consciously on their decisions and are thus thought to increase the degree to which participants' judgements reflect their implicit sensitivity to a given stimulus property (Schütze, 1996). Such tasks have been frequently used in monolingual sentence processing research to examine and compare participants' sensitivity to different types of linguistic information, or relative processing difficulty (compare e.g. McElree and Griffiths, 1995, 1998).

3.1 Participants

Twenty-two native Japanese-speaking learners of L2 English (20 female, two male) and 21 native speakers of English (nine female, 12 male; mean age: 25.0; SD: 8.3), all recruited from the University of Essex student community, participated in this experiment. None of the learners reported that they were fluent in any other foreign languages. All participants had normal or corrected-to-normal vision and had never been diagnosed as dyslexic. They received a small fee for their participation. The learners' first exposure to English was in a formal classroom setting at junior high school at the age of 12 or 13, and they had been living in the UK for around two years on average at the time of testing. Their general proficiency in English at the time of testing was assessed using the grammar part of the Oxford Placement Test (OPT; Allan, 1992). The learners' individual test scores indicated that their L2 grammatical proficiency ranged from 'mid-intermediate' to 'very advanced' on the OPT scale. Table 1 summarises the learners' bio-data and proficiency scores.

Table 1. *Summary of L2 learners' (n = 22) bio-data and proficiency scores; Experiment 1.*

	Mean	SD
Age (in years)	25.0	4.7
Age of first exposure	12.1	0.3
Length of immersion (in years)	1.9	2.1
OPT grammar score (in %)	78.0	6.9

To enable us to examine whether individual differences in working memory capacity would affect participants' performance in the speeded judgement task, reading span tests were conducted with both participant groups. The method was adopted from Daneman and Carpenter (1980) for the native speakers, and from Harrington and Sawyer (1992) for the L2 learners. The native speakers' group mean in the reading span test was 3.5 (SD: 1.1) out of a possible maximum of 6, and the L2 learners' group mean was 25.1 (SD: 5.0) out of a maximum of 42.[5]

All participants underwent both an untimed (Experiment 1A) and a speeded grammaticality judgement task (Experiment 1B). Prior to the start of the experiment, the learners were given a list of lexical items (verbs and nouns) to be used in the experiment and were asked to confirm whether they were indeed familiar with these items. Although described here first, the untimed task was administered after the speeded judgement task and a following break, so as to help ensure that prior exposure to the structures and lexical items used in the off-line task

[4] The fact that unlike Sturt (2003), Badecker and Straub (2002) found processing cost to be greater for a gender match between anaphor and inaccessible antecedent than for a gender mismatch could be due to use of different experimental techniques and/or materials.

[5] Harrington and Sawyer's (1992) L2 reading span test has been adapted for non-native speakers of English from Daneman and Carpenter's (1980) original test. Both tests involve reading through progressively larger sets of sentences followed by a word recall task, with Harrington & Sawyer's test using a slightly different scoring procedure from the original (total number of words recalled versus maximum set size for which words can be successfully recalled).

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

would not affect participants' performance in the main experiment.

3.2 Experiment 1A: untimed grammaticality judgements

Materials

The materials used in this experiment were canonically ordered declaratives containing finite complement clauses. The embedded verbs were all transitive and semantically compatible with a reflexive object. Locality violations involved reflexives that agreed in gender and number with the matrix subject but not with their local antecedent, as in (6a) below. In sentences involving c-command violations such as (6b), the reflexives agreed in gender with the non-c-commanding possessive modifier but not with the head noun of the local antecedent.[6]

(6) a. *John argued that the professors had criticised himself.
b. *The doctors believed that Mary's son had neglected herself.

The materials for the untimed judgement task comprised four sentences each from the two ungrammatical conditions and the same number of structurally parallel grammatical sentences. The 16 critical items were mixed with 16 grammatical and ungrammatical filler sentences of different structural types, and pseudo-randomised.

Procedure

All participants were tested individually in a quiet room. The test sentences were presented in the form of a written questionnaire. The participants were instructed to read each sentence carefully and to judge whether or not the sentence was well-formed and meaningful in English by ticking the corresponding "yes" or "no" box provided.[7] The order of correct answers was counterbalanced. There was no time limit in this task, which took approximately five to ten minutes to complete.

Results

Both native speakers (NSs) and learners (L2) performed highly accurately in this task, with overall mean accuracy rates of 96.1% (NSs) and 96.3% (L2). To correct for potential response biases as are frequently observed in binary forced-choice tasks, the participants' raw accuracy scores for the critical grammatical and ungrammatical items were converted to A′ scores (Pollack & Norman, 1964; Grier, 1971). By taking into account participants' judgements of both ungrammatical and grammatical items, A′ scores provide a unified sensitivity index to a given stimulus property, with an A′ score of .50 indicating chance performance and 1.00 perfect discrimination. Table 2 provides an overview of the two participant groups' mean accuracy rates per condition and their A′ scores.

Table 2. *Mean response accuracy (in percent) and A′ scores per condition per group (SDs in parentheses); Experiment 1A.*

		Locality	C-command
NSs ($n = 21$)	Grammatical	98.8 (5.46)	97.6 (7.52)
	Ungrammatical	94.0 (13.47)	94.0 (10.91)
	A′ scores	.98 (.03)	.97 (.05)
L2 ($n = 22$)	Grammatical	95.5 (9.87)	97.7 (7.36)
	Ungrammatical	95.5 (9.87)	96.6 (8.78)
	A′ scores	.98 (.04)	.98 (.04)

An ANOVA on participants' mean A′ scores with Sentence Type (locality vs. c-command) as a within-subjects factor and Group as a between-subjects factor showed no significant main effects and no interaction, suggesting that the learners were native-like in their ability to apply both the locality and the c-command requirements on reflexive binding.

3.3 Experiment 1B: speeded grammaticality judgements

Materials

The materials for the speeded grammaticality judgement task were structurally parallel to (but different from) those used in Experiment 1A. They included 32 grammatical and 32 ungrammatical sentences for each type of violation (local vs. c-command) and their grammatical controls, plus 16 filler sentences. The experimental items were distributed across two counterbalanced presentation lists, so as to avoid individual participants being exposed to both members of a given grammatical/ungrammatical pair.

The critical ungrammatical items all became ill-formed at the final word. Ungrammatical fillers involved embedded psych-stative verbs (e.g. *appreciate*) followed by a pronominal object and combined with an inanimate subject, resulting in a semantic violation that became apparent at the penultimate word. Including fillers of this kind was to help ensure that participants would read the stimulus sentences for meaning and to prevent them from focussing their attention exclusively on the sentence-final words. The experimental sentences were each divided into three presentation segments. Table 3 shows an example stimuli set, with the segmentation indicated by slashes.

[6] Complete lists of the experimental materials used in Experiments 1 and 2 are available upon request from the first author.

[7] Although experiments of this kind are likely to elicit acceptability rather than pure grammaticality judgements, we will follow others in using the term 'grammaticality judgement task' in a generic sense here.

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

Table 3. *Example stimuli used in the timed grammaticality judgement task; Experiment 1B.*

Condition	Grammatical	Ungrammatical
Locality	Mary believed / that the dancers / had hurt themselves.	Mary believed / that the dancers / had hurt herself.
C-command	The dancers believed / that Mary's brother / had hurt himself.	The dancers believed / that Mary's brother / had hurt herself.
Filler	Sally reported / that the actors / had fascinated her.	John confessed / that his sister's car / had appreciated him.

The number of singular and plural reflexives, and of masculine and feminine forms, was kept equal overall and counterbalanced across grammatical and ungrammatical conditions. The 64 experimental and 16 filler items in each presentation list were mixed with a further set of 80 structurally different filler sentences (half of which were ungrammatical) and pseudo-randomised.

Procedure

Participants were tested one at a time in a quiet laboratory setting. They were seated in front of a 17-inch PC monitor and were asked to read a set of English sentences presented to them in a segment-by-segment fashion, and to provide a grammaticality judgement at the end of each sentence. Each experimental trial started with a fixation cross being shown in the middle of the computer screen for 500 ms to indicate that stimuli presentation was about to begin. The stimuli were shown in white letters in Arial font at 26 point on a black background. Each segment remained on the screen for 1600 ms, then disappeared and was automatically replaced by the next.[8] The final segment was presented together with a full stop. Immediately after reading the final segment, the participants were required to judge, as quickly and as accurately as possible, whether the sentence was well-formed and meaningful in English by pressing either the "yes" or "no" button on a dual push-button box. The push-button box was activated from the onset of the final segment, and both response times and accuracy data were recorded. After every 20 sentences, stimulus presentation was halted and the instruction "Press 1 to see the next sentence" appeared on the screen, allowing the participants to take a break for as long as they needed.

The whole experimental session took approximately 40 minutes. The stimulus presentation and recording of the results were controlled by the NESU software package (Baumann, Nagengast and Klaas, 1993). Prior to the start of the experiment, a practice session was provided so that the participants could familiarise themselves with the procedure. When starting the practice session, the participants were reminded verbally by the experimenter that both speed and accuracy were important in this task. The practice session involved 12 practice sentences and two breaks.

Results

Prior to the statistical analyses of the data from the speeded judgement task, trials that exceeded the pre-determined timeout of 4500 ms for the NSs and 6000 ms for the learners were eliminated, which affected approximately 0.18% of the NSs' and 0.85% of learners' data.[9] The remaining data were then screened for outliers, and response times above or below 2.5 SDs from each participant's average RTs for each condition were also removed from the data set, affecting about 1.55% (NSs) and 0.63% (L2) of the data.

Participants' mean judgement accuracy was slightly lower overall than in the untimed task, with 94.1% (NSs) and 86.2% (L2) correct responses to grammatical items and 86.2% (NSs) and 68.3% (L2) correct responses to ungrammatical ones. As the differences between the grammatical and ungrammatical conditions are indicative of a bias towards "yes" responses in both groups, the participants' raw accuracy scores were converted to A′ scores. Table 4 summarises the native and non-native participant groups' mean accuracy and A′ scores per condition.

Note that compared to the results from the untimed judgement task (Experiment 1A), participants' ability to correctly reject locality violations dropped by 2.6% (NSs) and 21.8% (L2) in the speeded task, and their ability to reject c-command violations by 13.6% (NSs) and 17.7% (L2), respectively. A mixed ANOVA on participants' mean A′ scores revealed significant main effects of both Sentence Type ($F(1,41) = 11.084$, $p < .01$) and Group ($F(1,41) = 8.257$, $p < .01$) as well as a significant interaction between the two factors ($F(1,41) = 6.938$, $p < .05$). Paired sample t-tests showed that the learners

[8] The presentation time per segment was determined on the basis of pilot sessions with a group of Japanese speakers who did not participate any further in the current study, to ensure that the presentation time was as short as possible but long enough for the learners to be able to read the items properly.

[9] The timeouts were determined based on response times collected from pilot experiments with both native and non-native speakers.

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

Table 4. *Mean response accuracy (in percent) and A′ scores per condition per group (SDs in parentheses); Experiment 1B.*

		Locality	C-command
NSs ($n = 21$)	Grammatical	96.4 (4.2)	89.0 (8.8)
	Ungrammatical	91.4 (8.5)	80.4 (13.9)
	A′ scores	.97 (.03)	.90 (.07)
L2 ($n = 22$)	Grammatical	87.8 (12.9)	82.7 (10.2)
	Ungrammatical	73.7 (18.8)	78.7 (13.4)
	A′ scores	.88 (.10)	.87 (.09)

responded significantly less accurately than the native speakers to locality items ($t(41) = 4.084$, $p < .001$) but did not differ statistically from the former in their ability to judge c-command items ($t(41) = 1.311$, $p = .197$). The difference between the two sentence types was significant for the L1 group only ($t(20) = 4.833$, $p < .001$), with locality items being judged more accurately than c-command items.

Following common practice, statistical analyses of response times (RTs) were performed on correct trials only. Overall, grammatical strings were correctly accepted with mean latencies of 1308 ms by the native speakers and 2494 ms for the learners, and ungrammatical sentences were correctly rejected with mean latencies of 1456 ms (NSs) and 2506 ms (L2). Table 5 provides a summary of the two groups' mean RTs per condition.

As Table 5 shows, the learners took considerably longer to respond than the native speakers overall. Moreover, the two groups' RT patterns differed in that the native speakers judged locality items more quickly than c-command items, whereas the opposite was true for the learners. As the focus of this experiment was on participants' ability to detect violations of the locality and c-command constraints, a mixed ANOVA was carried out on the mean RTs to the two ungrammatical conditions. This analysis revealed a main effect of Sentence Type (with locality items eliciting longer RTs than c-command items overall) that was significant in the analysis by participants ($F_1(1,41) = 8.701$, $p < .01$; $F_2(1,62) = .738$, $p = .394$), a significant main effect of Group ($F_1(1,41) = 34.608$, $p < .001$; $F_2(1,62) = 149.823$, $p < .001$) and an interaction between Sentence Type and Group by participants that was marginal in the items analysis ($F_1(1,41) = 15.246$, $p < .001$; $F_2(1,62) = 2.897$, $p = .094$). Subsequent pair-wise comparisons revealed that only the learners took significantly longer, in the analysis by participants, to judge ungrammatical locality items than c-command items, whereas the between-conditions difference in the L1 group (which went in the opposite direction) did not prove statistically reliable (L1: $t_1(20) = -.984$, $p = .337$; $t_2(62) = -.783$, $p = .437$; L2: $t_1(21) = 3.986$, $p < .01$; $t_2(62) = 1.560$, $p = .124$).[10]

Effects of working memory

To investigate any potential effects of individual differences in working memory capacity (as measured by the L1 and L2 reading span tests) on participants' performance in the timed judgement task, separate per-group ANOVAs were conducted on A′ scores and RTs with the factor Sentence Type and with Reading Span as a covariate. No main effects or interactions with Reading Span were observed in the analyses of either of the two groups' A′ scores, indicating that the participants' ability to judge the experimental items was not affected by individual differences in their reading span scores. The analyses of the learners' RTs, on the other hand, revealed a significant main effect of Sentence Type ($F(1,20) = 11.847$, $p < .01$) and a significant interaction of Sentence Type and Reading Span ($F(1,20) = 7.717$, $p < .05$). The native speakers showed a main effect of Reading Span ($F(1,19) = 5.933$, $p < .05$) only. To further explore the source of the interaction observed in the L2 data, the learners were divided into two subgroups based on their median reading span score. The difference between the ungrammatical locality and c-command conditions was considerably larger for the low-span group (2848 ms vs. 2327 ms) than for the high-span group (2241 ms vs. 2114 ms). Statistically, the low-span participants proved typical of L2 group as a whole in that they responded significantly more slowly to locality than to c-command items ($t(10) = 4.331$, $p < .01$), whereas the difference between conditions did not reach statistical significance for the high-span subgroup ($t(10) = 1.708$, $p = .118$).

Effects of response speed on judgement accuracy

To examine whether either of the two participant groups had systematically sacrificed speed for accuracy (or vice versa) in the timed version of the task, we carried out correlation analyses between participants' individual mean RTs and their accuracy scores across all conditions. No significant correlations were found for either group, which means that there were no measurable overall speed–accuracy trade-off effects. To further test whether individual differences in response speed affected participants' judgement accuracy patterns across the experimental conditions, we carried out per-group ANOVAs on A′ scores with Sentence Type as a

[10] A parallel ANOVA for participants' RTs to grammatical items also yielded an interaction with Group ($F_1(1,41) = 6.375$, $p < .05$; $F_2(1,62) = 4.837$, $p < .05$), but the difference between conditions proved reliable only for the native speakers, who unlike the L2 group took longer to judge c-command than locality items, and was marginal by items ($t_1(21) = 3.432$, $p < .01$; $t_2(31) = 1.940$, $p = .057$).

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

Table 5. *Mean response latencies in milliseconds (SDs in parentheses) per group and condition; Experiment 1B.*

	Grammatical		Ungrammatical	
Group	Locality	C-command	Locality	C-command
NSs ($n = 21$)	1271 (432)	1365 (468)	1431 (428)	1476 (367)
L2 ($n = 22$)	2438 (591)	2327 (532)	2544 (694)	2220 (595)

within-subjects factor and participants' average overall response speed as a covariate. These analyses yielded no interactions with response speed, either.

3.4 Discussion

The results from Experiment 1 show that the learners' performance pattern across the experimental conditions differed from that of the native speakers in the speeded but not in the untimed version of the task. While the native speakers had more difficulty judging c-command than locality items, the learners' response times showed the opposite pattern, indicating that they had more difficulty identifying locality than c-command violations. Working memory capacity as measured by the two reading span tests did not affect the native speakers' judgement patterns and only had a small effect on the learners' performance such that the RT difference between the two sentence types (locality > c-command) was larger for learners with a relatively low reading span than for high-span learners. Participants' judgement accuracy, on the other hand, was affected neither by individual differences in reading span nor by their general processing speed (as reflected in their mean overall response speed).

The fact that the native speakers judged locality items with near-ceiling accuracy in both versions of the task is consistent with earlier findings suggesting that the locality requirement on reflexives immediately constrains antecedent search in L1 English (e.g. Nicol & Swinney, 1989). What might seem slightly puzzling, however, is that the native speakers found c-command items more difficult to judge than locality items in the speeded judgement task. One possible reason for this is the presence of a (structurally and conceptually) complex NP such as *Mary's brother* or *her brother's girlfriend* in the c-command items, which may have increased processing load relative to the otherwise structurally similar locality items (compare Walther, 1995).

Recall that unlike the native speakers, the learners were slowed down by the presence of a matching non-local antecedent that c-commanded the reflexive (compared to a non-c-commanding one) in the speeded judgement task, suggesting that they took the matrix subject into consideration as a potential binder before making their judgements. At the same time, this finding indicates that the learners' native-like sensitivity to the locality constraint on reflexives in the untimed judgement task did not merely reflect a 'processing' preference. If, in the untimed task, the learners had (correctly) focused on the reflexive's compatibility with the linearly closest potential antecedent only because such a locality strategy happens to be the more memory-friendly one, then this strategy should also have led to ceiling or near-ceiling judgement accuracy in the timed version of the task. Moreover, c-command items – where the binding-inaccessible competitor antecedent was linearly closer to the reflexive – should have been judged more slowly and/or less accurately than locality ones in this case, contrary to what we found.

4. Experiment 2: eye movements during reading

The results from Experiment 1 indicate that learners who prove indistinguishable from native speakers in an untimed judgement task may have difficulty applying the locality constraint on English reflexives in situations of increased processing demand. A number of follow-up questions arise from our first experiment, however, including the following:

- Can the observed asymmetry in learners' sensitivity to the locality and c-command requirements on reflexive binding be replicated in a more 'naturalistic' reading task?
- At what stage during processing is the non-local antecedent taken into consideration in L2 anaphor resolution?
- How do individual differences in reading span affect the real-time comprehension of reflexives?

In order to address these questions, we carried out a second experiment using eye-movement monitoring, a time-course sensitive technique that has also been shown to be suitable for studying non-native sentence processing (Frenck-Mestre, 2005a, b; Roberts, Gullberg and Indefrey, 2008). Recording participants' eye movements during reading should allow us to determine not only whether, but also when, during processing a binding-inaccessible antecedent is considered (compare Sturt, 2003).

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

Table 6. *Summary of L2 learners' (n = 32) bio-data and proficiency scores; Experiment 2.*

	Mean	SD
Age (in years)	27.1	7.6
Age of first exposure	12.3	1.2
Length of immersion (in years)	2.1	2.0
OPT grammar score (in %)	78.1	9.0

4.1 Participants

The L2 group consisted of 32 Japanese-speaking learners of English (27 female, five male) recruited from the University of Essex community. They were matched as closely as possible to the learners who took part in Experiment 1 with respect to their L2 grammar proficiency and English learning background. Table 6 shows the learners' OPT results, along with a summary of their relevant bio-data.

A group of 37 native speakers of English (17 female, 20 male; mean age: 23.7; SD: 7.7), served as a control group. They were also mostly from the student body of the University of Essex, and all had a university education. All participants had normal or corrected-to-normal vision and were naïve as to the ultimate purpose of the experiment. They were paid a small fee for their participation. To ensure that our learners had sufficient knowledge of the relevant binding constraints and were aware of the difference between reflexives and pronouns, all participants underwent an off-line multiple-choice antecedent identification task, details of which are described below. Furthermore, the two participant groups undertook the same reading span tests as in Experiment 1. The group means were 2.6 (SD: 0.8) out of a possible maximum of 6 for the native speakers and 27.8 (SD: 5.5) out of 42 for the learners.

The native speakers were tested in a single session, with the on-line experiment (Experiment 2B) being administered before the off-line one (Experiment 2A) so that participants would be fully alert for the on-line task. The L2 learners were tested in two separate sessions, with at least a week in between the main experiment and the corresponding off-line task.

4.2 Experiment 2A: multiple choice antecedent identification

Materials

The materials for the off-line antecedent identification task comprised a total of 60 sentences which were designed to probe participants' knowledge of binding principles A, B and C. In total, there were ten types of sentence (six of each type), manipulating features such as the gender of the binding-inaccessible antecedent, its relative hierarchical prominence (i.e. c-command), and the type of pronoun. Of particular interest to the current study are the three types of sentence that corresponded to the materials used in the on-line task (described further below), examples of which are given in (7a–c).

(7) a. *Inaccessible match, c-command*

Adam believes that Ian blames himself.

b. *Inaccessible mismatch, c-command*

Diana knows that Mark helped himself.

c. *Inaccessible match, no c-command*

It appears to Maria that Jane cut herself.

To examine whether our learners were aware of the difference between reflexives and pronouns in English, the materials for the off-line task also included sentences that tested participants' knowledge of Principle B, which requires that pronouns be free in their local domain (Chomsky, 1981), such as those shown in (8a) and (8b) below, with the inaccessible (local) antecedent either matching the pronoun in gender or not.

(8) a. *Inaccessible match*

Daniel recalled that Richard had woken him.

b. *Inaccessible mismatch*

Lucy says that Adam surprised her.

Each sentence was followed by a grid offering four answer choices (including "Perfect", "Possible", "Not possible" and "Don't know") for different potential antecedents. The experimental items were randomised twice to create two presentation lists with the items appearing in different order, and half of the participants in each language group saw each list.

Procedure

The participants were tested individually in a quiet laboratory room. The antecedent identification task was administered as a written questionnaire. Participants were asked to decide whether each of several referents was a "perfect", "possible" or "impossible" match for the (underlined) reflexive or pronoun in each sentence. Participants were set no time limit for this task, which took them, on average, around 20 minutes to complete.

Results

The purpose of the multiple-choice antecedent-identification task was to establish whether the non-native participants' ultimate antecedent choices were native-like and in line with Principle A, and whether they were aware of the difference between reflexives and pronouns. We thus analysed participants' responses to both the Principle A and the Principle B items. The learners' response patterns

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

Table 7. *Overview of native speakers' and L2 learners' off-line antecedent choices (in percent, SDs in parentheses) for the three critical Principle A conditions; Experiment 2A.*

	NP1 impossible		NP1 possible		NP2 perfect	
	NSs	L2	NSs	L2	NSs	L2
Inaccessible match, c-command	84.2 (31.8)	86.5 (25.6)	8.8 (21.0)	10.4 (21.9)	91.3 (21.7)	96.9 (7.8)
Inaccessible mismatch, c-command	90.4 (26.9)	97.4 (7.5)	3.8 (11.6)	1.6 (6.5)	94.2 (17.5)	98.4 (4.9)
Inaccessible match, no c-command	86.7 (28.5)	84.9 (26.2)	8.3 (20.7)	13.0 (22.7)	92.5 (19.2)	96.4 (8.2)

were very similar to the native speakers' overall. For sentences testing their knowledge of reflexive binding ($n = 30$), our participants correctly identified binding-inaccessible noun phrases as 'impossible' antecedents the vast majority of the time (NSs: 86.7%; L2: 85.3%) while considering the accessible antecedent to be a 'perfect' match for the reflexive in almost all cases (NSs: 91.3%; L2: 95%). However, both groups also occasionally considered the binding-inaccessible antecedent 'possible' (NSs: 7.6%; L2: 10.8%).[11] Table 7 presents an overview of participants' antecedent choices for the three critical Principle A conditions (7a–c). Mann-Whitney tests confirmed that the two groups' proportions of antecedent choices did not differ significantly for any of these.

For Principle B sentences ($n = 18$), the learners also performed at or above the native speakers' accuracy level, with the NSs correctly ruling out the local NP as a potential binder for a pronoun 87.5% (SD: 21.6) and the L2 group 92.8% (SD: 10.7) of the time overall. Taken together, the results from the questionnaire task confirm that the learners were aware of the different binding requirements of reflexives and pronouns in English.

4.3 Experiment 2B: eye-movement monitoring

Materials

For the eye-movement monitoring experiment, 24 sets of short texts were prepared consisting of a lead-in sentence, a second (critical) sentence that contained the reflexive and a 'closing' sentence. The experiment had a 2 × 2 design with each presentation list containing six items for each of the four experimental conditions. The critical sentences were similar to those used by Sturt (2003) but modified so as to avoid using gender stereotype nouns, because of possible cross-cultural differences in the stereotypical gender assigned to them. Instead, the binding-inaccessible antecedent's gender was manipulated using a set of common English male and female proper names. To minimise the possibility that these names might not be recognised by L2 learners as being associated with a specific gender, a small pre-test was carried out. Fifty names from the *Oxford Learner's Dictionary* (Hornby, 2000) were given to five Japanese speakers who were asked to rate on a scale of 1–7 how 'male' or 'female' they considered them to be. Only names that were rated 1 ("definitely male") or 7 ("definitely female") were included in the materials for this experiment, with each alternating pair (e.g. *John*/*Jane*) matched for length. There was no manipulation of the accessible antecedent's gender, so that each sentence contained a syntactically and pragmatically licit antecedent for the reflexive pronoun. Each sentence quadruplet consisted of two c-command and two non-c-command conditions as shown in (9).

(9) John/Jane and Richard were very worried in the kitchen of the expensive restaurant.

a. *Inaccessible match, c-command*

John noticed that Richard had cut himself with a very sharp knife.

b. *Inaccessible mismatch, c-command*

Jane noticed that Richard had cut himself with a very sharp knife.

c. *Inaccessible match, no c-command*

It was clear to John that Richard had cut himself with a very sharp knife.

d. *Inaccessible mismatch, no c-command*

It was clear to Jane that Richard had cut himself with a very sharp knife.

Kitchens can be dangerous places.

In the lead-in sentence, two named characters are introduced that are then mentioned again in the next sentence. However, only one of these is a legitimate binder for the reflexive. In all of (9a–d) above, for instance,

[11] Although the overall between-groups difference in "NP1 possible" responses proved to be statistically reliable ($U = 472$, $p < 0.05$), separate per-condition comparisons indicated that this was largely due to the learners' relatively high proportion (20.3%) of "NP1 possible" choices in the bound variable condition (e.g. *Andrew recalled that every man examined himself*), which represents semantic binding without coreference (compare e.g. Büring, 2005, pp. 81–92).

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

Richard is the sole binding-accessible antecedent for *himself*. In (9a), the binding-inaccessible antecedent (i.e. the matrix subject *John*) matches the reflexive in gender and also c-commands it, whereas in (9c) the binding-inaccessible antecedent also matches the reflexive but does not c-command it.

Following Sturt's (2003) finding that Principle A may be violable at a later point in processing, we expect the native speakers to be unaffected by the inaccessible antecedent's gender during earlier processing stages, with the possibility that such effects might be seen in later measures. If initial antecedent search in the L2 is constrained by Principle A in the same way as in native speakers and binding-inaccessible antecedents are considered only as an 'afterthought', then effects of the above gender manipulation should also be restricted to later measures in the L2 group. If, on the other hand, initial antecedent search is affected by a noun phrase's relative structural or discourse prominence rather than by binding constraints, then effects of the inaccessible antecedent's gender might in fact be seen from relatively early on during parsing.

In addition to the 24 experimental texts, 48 filler texts were also constructed which differed structurally from the experimental items and involved various other types of noun-pronoun coreference relationships. To ensure that the participants would read the stimuli texts properly and for meaning, *yes/no* content questions followed all the experimental and half of the filler items. The questions for the experimental sentences never asked about any aspect of the binding relationship between reflexive and antecedent, in order to keep the participants' focus away from the experimental objective. Four randomised presentation lists were constructed using a Latin square design.

Procedure

The eye-movement monitoring experiment was run using a head-mounted SR Research EyeLink II system, which has a sample resolution of 500 Hz. Although reading was binocular, tracking was monocular, choosing the eye that had the best calibration. A screen of instructions was presented, informing participants of the task, and then a short trial session was run. At the beginning of each trial a black dot was displayed on a white background at the top left of the screen, which the participant had to fixate on and then press the "A" button on a gamepad. This had the twofold purpose of correcting any drift in the eye tracker before each trial and having the participant's eyes in the correct place to read the first word of the upcoming text, which appeared as soon as the button was pressed in the same place as the dot. The stimuli were presented on several lines in a fixed-width font (Courier New) at 21 point, with line breaks inserted such that the reflexive would appear roughly in the centre of the screen in each critical trial. An example of how the stimuli texts were presented is shown in (10) below (with line breaks indicated by hash marks).

(10) John and Richard were very worried in the kitchen of the expensive restaurant. #

John noticed that Richard had cut himself with a very sharp knife. #

Kitchens can be dangerous places.

The participants were asked to read through each text in a normal manner and press the "A" button when they had finished. On trials where a question followed, they also had to press either the "A" button for "yes" or the "B" button for "no". There were two enforced breaks during the experiment, but participants were informed that they were free to take a break after any trial. For the L1 group the whole procedure took between 25 and 40 minutes, and between 35 and 55 minutes for the L2 group.

Results

Both participant groups answered the comprehension questions following the experimental items correctly 90% or more of the time in all experimental conditions, indicating that they understood the sentences and were paying attention to the task. The native group and the learners showed a mean overall response accuracy of 95% and 92%, respectively.

The eye-movement data set from one native speaker was eliminated from the analysis because it was incomplete. Prior to the analysis of the remaining 68 participants' reading time data, fixations of less than 50 ms were automatically merged with the nearest fixation if it was within 1° of visual arc. Participant blinks and tracker inaccuracies led to the loss of 0.9% of the NSs' data and 1.3% of the learners' data. Moreover, if a given region of interest was not fixated at all before entering a subsequent region, the data for that trial was excluded from further analysis, leading to the elimination of 18.3% of the NSs' data and 2.6% of the L2 data. Finally, a small number of extreme outlier data points were removed which accounted for 0.6% of the remaining native data and 0.6% of the L2 data.[12]

The experimental sentences were divided into different regions of interest. Statistical analyses were carried out for reading times in the reflexive (*himself/herself*), post-critical (the two words following the reflexive) and final (the remainder of the sentence) regions. For each of these regions, the following eye-movement measures were

[12] Extreme outliers were identified by the SPSS statistical analysis software package, with individual data points three times or more the interquartile range (IQR) above the third quartile being eliminated from the data set. IQR was calculated per condition but collapsed over subjects, separately for each group.

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

Table 8. *Reading times in the reflexive region in milliseconds (SDs in parentheses) and percentages of "regressions in" for the inaccessible antecedent; Experiment 2B.*

		First fixations	First pass	Regression path	Second pass	Total reading time	Regressions in %
Inaccessible match,	NS	224 (42)	238 (51)	307 (127)	129 (131)	364 (142)	79
c-command	L2	283 (52)	419 (115)	517 (158)	285 (215)	701 (264)	70
Inaccessible mismatch,	NS	218 (43)	234 (54)	298 (108)	111 (86)	345 (104)	71
c-command	L2	294 (55)	372 (88)	445 (145)	269 (186)	655 (218)	65
Inaccessible match,	NS	216 (43)	233 (45)	277 (84)	106 (117)	330 (109)	34
no c-command	L2	285 (52)	370 (91)	471 (148)	295 (199)	679 (237)	30
Inaccessible mismatch,	NS	214 (35)	230 (47)	270 (92)	126 (107)	342 (107)	35
no c-command	L2	285 (61)	393 (95)	457 (127)	252 (163)	651 (198)	29

Table 9. *Reading times in the postcritical region in milliseconds (SDs in parentheses); Experiment 2B.*

		First pass	Regression path	Second pass	Total reading time
Inaccessible match,	NS	301 (90)	400 (146)	113 (97)	423 (135)
c-command	L2	438 (134)	527 (188)	215 (170)	665 (235)
Inaccessible mismatch,	NS	290 (95)	389 (148)	108 (81)	410 (140)
c-command	L2	496 (139)	535 (162)	249 (168)	751 (205)
Inaccessible match,	NS	300 (103)	371 (141)	110 (124)	417 (157)
no c-command	L2	472 (105)	537 (179)	242 (142)	718 (217)
Inaccessible mismatch,	NS	307 (73)	415 (150)	101 (108)	424 (155)
no c-command	L2	485 (115)	569 (186)	238 (175)	744 (258)

analysed: (i) FIRST FIXATION DURATION, the length of participants' initial fixation on a region; (ii) FIRST-PASS DURATION, the sum of all fixations in a region before the reader exits the region in either direction; (iii) REGRESSION PATH DURATION, the sum of all fixations from the time when the region is first entered until it is first exited to the right; (iv) SECOND-PASS DURATION, which is calculated by subtracting the first-pass duration from the (v) TOTAL READING TIME, the sum of all fixations in a region.[13] Another measure of potential interest here is REGRESSIONS IN, which refers to the proportion of trials in which a region was read again after the readers' eyes had already moved away from it to the right. In the case of the current experiment, the critical region to look at here would be the inaccessible antecedent, where more 'regressions in' could be taken to suggest that readers evaluated its suitability as an antecedent for the reflexive.

[13] The post-critical and final sentence regions were analysed so as to check for possible 'spillover' or delayed effects. Recall that in Sturt's (2003) study, for example, manipulating the binding-inaccessible antecedent's gender only affected reading times in the pre-final sentence region.

To determine whether the native and non-native participants processed the reflexive and following regions differently, mixed ANOVAs were carried out for each measure with the factors Inaccessible Antecedent (match vs. mismatch), C-command (yes/no) and Group (NSs vs. L2). Tables 8–10 provide an overview of the two participant groups' mean reading times on the reflexive, postcritical and final regions per condition. Table 8 additionally shows the proportion of 'regression in' to the inaccessible antecedent upon participants' encountering of the reflexive. Main effects of Group were found for all measures and across all three regions of interest, reflecting the fact that the learners read the experimental stimuli more slowly than the native speakers overall. Other significant main effects or interactions were observed for the reflexive region only.

While the analysis of first fixation durations yielded no main effects or interactions, we found a significant interaction between the factors Inaccessible Antecedent and C-command ($F_1(1,66) = 7.944, p < .01$; $F_2(1,23) = 7.272, p < .05$) for first-pass durations, as well as a three-way interaction with the factor Group ($F_1(1,66) = 7.798$, $p < .01$; $F_2(1,23) = 5.249, p < .05$). Subsequent pair-wise comparisons showed that the learners took significantly

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

Table 10. *Reading times in the final region in milliseconds (SDs in parentheses); Experiment 2B.*

		First pass	Regression path	Second pass	Total reading time
Inaccessible match,	NS	326 (121)	649 (358)	77 (86)	421 (168)
c-command	L2	615 (164)	1136 (592)	223 (220)	858 (318)
Inaccessible mismatch,	NS	290 (119)	617 (324)	89 (88)	393 (176)
c-command	L2	605 (212)	1283 (580)	224 (199)	842 (308)
Inaccessible match,	NS	288 (121)	626 (374)	89 (110)	397 (179)
no c-command	L2	599 (164)	1240 (482)	196 (161)	808 (237)
Inaccessible mismatch,	NS	299 (106)	658 (355)	84 (120)	385 (172)
no c-command	L2	605 (197)	1225 (609)	199 (150)	806 (285)

longer to read the reflexive when a c-commanding but binding-inaccessible antecedent matched the reflexive's gender, as in (9a) above, than when it did not match the reflexive in gender, as in (9b) ($t_1(31) = 2.578, p < .05$; $t_2(23) = 2.711, p < .05$). No such 'gender-match' effect was found for the two non-c-command conditions ($t_1(31) = 1.413$, $p = .168$; $t_2(23) = 1.329$, $p = .197$), however. The native speakers, in contrast, did not show any statistical differences between conditions in their first-pass reading times (c-command pair: $t_1(35) = .467$ $p = .644$; $t_2(23) = .386$ $p = .703$; non-c-command pair: $t_1(35) = .379$, $p = .707$; $t_2(23) = .586$, $p = .563$). For regression path times, the numerical pattern was very similar to that seen in the first-pass times. A mixed ANOVA showed main effects of Inaccessible Antecedent ($F_1(1,66) = 4.890, p < .05$; $F_2(1,23) = 6.263, p < .05$) and C-command, the latter failing to reach significance in the subjects analysis ($F_1(1,66) = 3.503, p = .066$; $F_2(1,23) = 4.572, p < .05$), as well as a marginal three-way interaction with Group in the analysis by items ($F_1(1,66) = 1.492$, $p = .226$; $F_2(1,23) = 4.300$, $p = .05$). Note that the main effect of Inaccessible Antecedent, which reflects the fact that overall, the processing of the reflexive was slowed down by the availability of a gender-matched but binding-inaccessible antecedent, was carried mainly by the L2 group. While the 'match' conditions only showed small numerical disadvantages of 9 ms (for c-commanding inaccessible antecedents) and 7 ms (for non-c-commanding inaccessible antecedents) compared to the 'mismatch' conditions in the native speakers' data, the L2 learners were slowed down, on average, by 72 ms in the c-command condition and by 14 ms in the non-c-command condition.

No significant effects or interactions were seen in the analyses of second-pass durations or total reading time, but there was a highly significant main effect of the factor C-command in the analysis of 'regressions in' to the inaccessible antecedent ($F_1(1,66) = 131.847$, $p < .001$; $F_2(1,23) = 317.854$, $p < .001$), which shows that both groups re-read the inaccessible antecedent far more often when it c-commanded the reflexive than when it did not, though there are other differences between conditions that could also explain this effect (see below for more discussion on this issue). The main effect of Group, reflecting the overall slightly smaller proportion of regressions to the inaccessible antecedent in the L2 group, was significant by items only ($F_1(1,66) = 1.900$, $p = .173$; $F_2(1,23) = 8.068$, $p < .01$). Although the proportion of looks back to the inaccessible antecedent was numerically higher in the gender-match condition (9a) than in the corresponding non-match condition (9b), there was no statistically significant effect of the inaccessible antecedent's gender and no interactions.

The ANOVAs for the postcritical region only showed some marginal main effects of Inaccessible Antecedent in first-pass durations ($F_1(1,66) = 2.763$, $p = .101$; $F_2(1,23) = 3.386$, $p = .079$) and total reading times ($F_1(1,66) = 2.963$, $p = .09$; $F_2(1,23) = 2.755$, $p = .111$), and marginal interactions of the factors Inaccessible Antecedent and Group in the subjects analyses of first-pass durations ($F_1(1,66) = 3.521$, $p = .065$; $F_2(1,23) = 1.673$, $p = .209$) and total reading times ($F_1(1,66) = 3.678$, $p = .059$; $F_2(1,23) = 1.382$, $p = .252$). No significant or marginal effects (other than main effects of Group) or interactions were observed in the reading times in the final region.

To determine whether participants' reading patterns were influenced by individual differences in their reading span, we carried out additional per-group ANOVAs with the factor Reading Span as a covariate. We found no significant interactions with Reading Span for any of the experimental measures in either of the two participant groups, however.

4.4 Summary

Summarising, we saw that the initial processing of reflexives by Japanese learners – but not by the native speaker controls – was influenced by a structurally and discourse-prominent but binding-inaccessible antecedent.

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

Note that unlike Sturt (2003), we found no reliable effects of the inaccessible antecedent's gender in any of the later eye-movement measures or sentence regions in either participant group, which is likely to be due to differences between the design and/or the materials used in the two studies. Recall that our materials differed from Sturt's in that we avoided using gender-stereotyped nouns and only manipulated the gender of the binding-inaccessible antecedent. Moreover, in our experiment both potential antecedents were initially introduced in the lead-in sentence as conjoined subject noun phrases, which may have reduced the relative discourse-prominence of the inaccessible versus the accessible antecedent, compared to Sturt's experiment. The larger proportions of 'regressions in' to the inaccessible antecedent in the c-command conditions could be taken to suggest that the native speakers, too, considered a c-commanding non-local NP at some (later) stage during processing. However, as has been pointed out to us by a reviewer, this finding should be interpreted with caution as this effect is confounded with the fact that the inaccessible antecedent was presented at the start of the sentence (and line) in the c-command conditions and several words into the sentence in the non-c-command conditions. The results from our eye-movement experiment will be further discussed below, together with the results from Experiment 1.

5. General discussion

The current study sought to gather (convergent) evidence from different experimental tasks on non-native speakers' sensitivity to binding constraints in L2 processing. Experiment 1 examined learners' metalinguistic grammatical abilities using judgement tasks, whereas Experiment 2 investigated learners' interpretation of reflexives using comprehension tasks. Taken together, our results show that our Japanese-speaking participants had difficulty applying binding Principle A in on-line processing tasks but not in corresponding untimed tasks. Specifically, we found that our learners were slowed down by the presence of a matching but binding-inaccessible antecedent that c-commanded the reflexive both in a speeded metalinguistic judgement task and in an on-line reading comprehension task that was more naturalistic than the former in that the stimuli sentences were presented uninterrupted and participants were free to read them at their own speed. The results from Experiment 1 showed that the learners' sensitivity to the locality requirement on reflexive binding was more vulnerable to processing pressure than the native speakers', and Experiment 2 revealed that the learners considered the matrix subject as a potential antecedent for an embedded reflexive as soon as this was encountered. That effects of a c-commanding but binding-inaccessible antecedent were restricted to first-pass reading times indicates that the learners were confused during early processing stages only, which would fit with our observation that their ultimate antecedent choices were in fact native-like. In what follows, we will discuss possible factors that may have influenced the learners' performance in turn.

5.1 Processing factors

As is usually the case in L2 processing studies, our learners showed longer response latencies (Experiment 1) and reading times (Experiment 2) than the native speakers overall, which confirms that reading in a non-native language is cognitively more demanding than native reading, especially when reading in a foreign script (compare e.g. Akamatsu, 1999; Koda, 2005). Recall that with regard to reflexive binding, it is conceivable that learners of English who appear to adhere to Principle A in off-line judgement or comprehension tasks may in fact choose, or focus on, the local antecedent simply because this is the most memory-friendly option. The results from the current study, however, suggest that this is not the case. Instead, we found that contrary to what would have been expected if they had followed a simple locality strategy, our learners were measurably affected by the presence of a matching non-local antecedent in situations of increased processing demand only, both in the timed judgement task and during real-time anaphor resolution. Although the learners were generally slower than the NSs, participants' reading times were not modulated by individual differences in reading span in the eye-movement monitoring task, where they were able to re-read earlier parts of the stimuli texts. Thus, it is likely that the small reading-span effect observed in Experiment 1 was due to the way the experimental stimuli were presented, with low-span learners having more difficulty retrieving the matrix subject from memory than their high-span peers and/or suppressing the reactivation of a matching matrix subject when encountering a matching reflexive. The limited influence of (verbal) working memory capacity as measured by standard L1 and L2 reading span tests is consistent with previous L2 sentence processing studies that have found no or limited working memory effects (Juffs, 2004, 2005; Felser and Roberts, 2007). Recall further that participants' performance patterns across the critical conditions in Experiment 1 were not influenced by individual differences in processing speed (as reflected in their individual average response speed). In short, it appears that the observed L1/L2 differences in the processing of reflexives cannot simply be attributed to domain-general 'processing' factors such as cognitive resource limitations, or to slower processing speed.

5.2 L1 influence

As noted earlier, the closest translation equivalent to English reflexives in Japanese is the polymorphemic anaphor *kare*/*kanojo-zisin* "he/she-self", which, unlike

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

the primary reflexive *zibun*, is gender-marked and normally requires local binding. It seems reasonable to assume, then, that our Japanese participants were aware of the differences in binding requirements between mono- and polymorphemic reflexives, at least in their native language. The results from the off-line antecedent identification task in Experiment 2 moreover demonstrate that they were also able to differentiate between reflexives and pronouns in English and did not differ from native speakers in their ability to apply the relevant binding constraints in their L2.

However, given that the more commonly used reflexive *zibun* permits LD binding, it is conceivable that our learners might have transferred the LD binding property of *zibun* to English reflexives during processing. Unlike the locality requirement, which is peculiar to English-type argument reflexives, the c-command requirement on reflexive binding has been argued to hold universally (compare Büring, 2005), which could explain why our learners had comparatively less difficulty applying the latter. Although the fact that we investigated learners from a single L1 background only does not allow us to draw any strong or definite conclusions about the possible role of L1 background on L2 anaphor resolution, we do not think that transfer from Japanese provides a particularly convincing explanation for the observed L1/L2 processing differences, however. Note that the fact that our learners demonstrated native-like knowledge of Principle A in the untimed tasks raises the question of why the transfer of LD binding should have been restricted to on-line processing tasks. A possible answer to this is that influence from the native language may be more difficult to suppress when processing demands are high (compare Frenck-Mestre, 2005b). Even if we take this hypothesis to be correct, however, it is not immediately obvious which property or properties of Japanese might have been subject to transfer here. Since Japanese differs from English in, among other things, being a head-final language, any direct word-by-word or phrase-by-phrase mapping of Japanese LD-binding configurations (as, for example, in (1b) above) onto our English stimuli sentences would have been impossible.

A more realistic possibility would thus be lexically-based transfer, with the learners transferring the morphosyntactic properties of *zibun* to English reflexives, in effect misanalysing the latter as gender-neutral, morphologically simplex anaphors that do not require local binding. This is unlikely not only because of the availability of gender-marked, polymorphemic anaphors in Japanese, but also because in Experiment 2B, the Japanese participants did in fact demonstrate sensitivity to the reflexives' gender features during their initial processing of the reflexive region.

Considering the above observations, the idea that our learners should have transferred the morphosyntactic properties of *zibun* to English reflexives during processing seems rather implausible. An alternative possibility that we cannot rule out entirely, however, is that of transfer occurring not at the lexical but at the discourse-pragmatic level, with the learners misinterpreting English argument reflexives as logophors rather than structural anaphors despite being aware of the reflexives' morphosyntactic feature.[14] According to Reinhart & Reuland (1993), logophors are discourse-bound rather than syntactically bound. However, even though English reflexives are also sometimes used logophorically, as in *The recent picture of herself in the newspaper upset Mary*, note that our experimental sentences did not provide a typical logophoric environment. The 'logophor' hypothesis also raises the question of why our learners should then have abandoned this interpretation again during later processing stages – a question that we will return to below.

In summary, even though previous studies examining the processing of syntactic ('filler–gap') or referential dependencies in the L2 have found no (Williams, Möbius and Kim, 2001; Marinis et al., 2005; Roberts et al., 2008) or little (Juffs, 2005) evidence of L1 influence on learners' processing performance, the possible role of language background in learners' processing of reflexives should nevertheless be investigated more systematically, which we plan to do in future research.

5.3 Reduced syntactic processing

Given the results from earlier L2 processing studies indicating that even highly proficient L2 learners have more difficulty than native speakers establishing syntactic dependencies in real time (see Clahsen and Felser, 2006, for a review), another possibility of accounting for our results is the hypothesis that learners rely more on semantic or pragmatic cues to the interpretation of reflexives. Information provided by the pragmatic context, for example, has been shown to influence L2 anaphor resolution in both off-line (Demirci, 2000; Lee, 2008) and real-time processing tasks (Roberts et al., 2008).

Comparing Turkish-speaking learners' interpretation of English sentences such as (11a), which were pragmatically biased towards the non-local antecedent (*the little boy*), and those that were biased towards the local antecedent (*the famous actor*), such as (11b), Demirci (2000), for example, found that the learners' antecedent choices were much more strongly influenced by pragmatic bias than native speakers' in both directions, and across different proficiency levels.

(11) a. The little boy was angry that his father always hit himself.

b. Ali heard that the famous actor talked about himself on TV.

[14] For some discussion of the conditions for using Japanese *zibun* logophorically, see Oshima (2007) and references cited therein.

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

For finite NP1-biased sentences, Demirci's learner groups selected the non-local NP between 75% and 95% of the time on average (native speakers: 19%), whereas for NP2 biased ones they correctly chose the local NP in 82%–92% of cases (NSs: 100%). Demirci's results thus suggest that pragmatic information can override syntactic constraints in L2 – but not L1 – anaphor resolution.

Examining on-line pronoun resolution in L2 Dutch using eye-movement monitoring, Roberts et al. (2008) found that highly proficient Turkish- and German-speaking learners' reading times would slow down when there were two – as opposed to only one – potential antecedents (i.e. *Peter* and *Hans*) present in the discourse for ambiguous subject pronouns such as *hij* "he" in (12) below.

(12) Peter en Hans zitten in het kantoor. Terwijl Peter aan het werk is, eet hij een sandwich.

"Peter and Hans are in the office. While Peter is working, he is eating a sandwich."

This kind of referential ambiguity did not cause any slow-down at all in the native speaker controls, however.

Our results are consistent with Roberts et al.'s findings in that the presence of a matching discourse-prominent 'competitor' antecedent also increased processing difficulty for our learners, but not for the native speaker controls. The results from our second experiment furthermore revealed interesting differences between the time-course of L1 and L2 processing, showing that the learners considered the matrix subject during their initial processing of the reflexive rather than as an afterthought, as has been reported for English native speakers in a similar eye-movement study (Sturt, 2003). This suggests that Principle A does not act as an initial filter constraining antecedent search in L2 English.

Our eye-movement results indicate that the learners were confused by a matching non-local antecedent only temporarily, though. There was no influence of the inaccessible antecedent's gender on second-pass or total reading times in the critical region, or in any measure in later regions, and the learners' ultimate (off-line) antecedent choices were also native-like. The first-pass reading-time pattern seen in our Japanese participants, compared to the absence of any such effects in the native speakers' reading times, would be expected if the initial processing of reflexives were influenced more strongly by discourse-level or pragmatic cues in non-native than in native sentence comprehension. Recall that our c-command manipulation effectively served to alter the inaccessible antecedent's relative discourse prominence, with competitor antecedents in matrix subject position being more salient than those embedded more deeply within a prepositional phrase (as in *It was clear to John*. . .).

In short, while we cannot completely rule out some degree of L1 transfer at the pragmatic level, our results are consistent with the assumption that L2 learners over-use discourse-pragmatic information during on-line anaphor resolution compared to native speakers. Although an increased reliance on pragmatic cues in on-line processing need not necessarily compromise learners' ability to successfully comprehend the L2 input, it may put them at a disadvantage, compared to native speakers, by temporarily increasing the degree of (perceived) ambiguity in on-line reference resolution. The observation that in untimed tasks, learners' ultimate antecedent choices may nevertheless be native-like might be indicative of a greater reliance on explicit or 'declarative' (as opposed to implicit or 'procedural') grammatical knowledge in L2 sentence comprehension, the recruitment of which may be delayed.[15] Clearly though, the relative timing of syntactic versus pragmatic information in L2 processing, and how this compares with the time-course of native language processing, requires further systematic investigation.

6. Concluding remarks

The results from two qualitatively different chronometric tasks – a speeded judgement task and an on-line reading comprehension task – both indicated that our learners had more problems than native speakers applying the locality constraint on English reflexive binding during processing. The fact that this difficulty was not discernible in the learners' performance in corresponding untimed tasks highlights the importance of language processing data for helping us gain a more comprehensive picture of non-native speakers' linguistic abilities. The results from the eye-movement monitoring task, in particular, provided detailed information about the time-course of L2 reflexive anaphor resolution which showed that unlike what has been reported in a similar monolingual eye-movement study (Sturt, 2003), L2 learners violate the locality requirement of Principle A during early processing stages – even though their ultimate antecedent choices may be native-like. Our results further indicate that 'processing' factors such as working memory limitations or slower processing speed cannot fully account for the observed L1/L2 differences. Instead, our results suggest that L2 learners are more strongly influenced than native speakers by semantic or pragmatic cues such as gender congruence or discourse salience when interpreting reflexive anaphors, which may serve to compensate for a reduced ability to apply grammatical coreference constraints when processing the L2 input. The question of how different information sources interact

[15] See Ullman (2005) for some discussion of the possible neurophysiological basis for L1/L2 differences with respect to automatic structural processing.

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

during L2 sentence comprehension clearly requires further investigation, however, as does the extent to which L2 learners' processing of anaphoric expressions may be influenced by properties of their L1.

References

Akamatsu, N. (1999). The effects of first language orthographic features on word recognition processing in English as a second language. *Reading and Writing*, 11, 381–403.

Allan, D. (1992). *The Oxford placement test*. Oxford: Oxford University Press.

Badecker, W. & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 748–769.

Baumann, H., Nagengast, J. & Klaas, G. (1993). New Experimental Setup (NESU). Ms., Max Planck Institute for Psycholinguistics, Nijmegen.

Büring, D. (2005). *Binding theory*. Cambridge: Cambridge University Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

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

Cole, P. & Sung, L.-M. (1994). Head movement and long-distance reflexives. *Linguistic Inquiry*, 25, 355–406.

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

Demirci, M. (2000). The role of pragmatics in reflexive interpretation by Turkish learners of English. *Second Language Research*, 16, 325–353.

Eckman, F. (1994). Local and long-distance anaphora in second-language acquisition. In E. Tarone, S. Gass & A. Cohen (eds.), *Research methodology in second-language acquisition*, pp. 207–225. Hillsdale, NJ: Lawrence Erlbaum.

Felser, C. & Roberts, L. (2007). Processing *wh*-dependencies in a second language: A cross-modal priming study. *Second Language Research*, 23, 9–36.

Frenck-Mestre, C. (2005a). 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.

Frenck-Mestre, C. (2005b). Ambiguities and anomalies: What can eye-movements and event-related potentials reveal about second language sentence processing? In J. Kroll & A. M. B. de Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 268–281. Oxford: Oxford University Press.

Grier, J. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 75, 424–429.

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

Harris, T., Wexler, K. & Holcomb, P. (2000). An ERP investigation of binding and coreference. *Brain and Language*, 75, 313–346.

Hawkins, R. (2001). *Second language syntax: A generative introduction*. Oxford: Blackwell.

Hirakawa, M. (1990). A study of the L2 acquisition of English reflexives. *Second Language Research*, 6, 60–85.

Hornby, A. S. (2000). *Oxford advanced learners dictionary of current English* (6th edn.). Oxford: Oxford University Press.

Juffs, A. (2004). Representation, processing, and working memory in a second language. *Transactions of the Philological Society*, 102, 199–225.

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.

Koda, K. (2005). *Insights into second language reading*. Cambridge: Cambridge University Press.

Lee, K.-Y. (2008). The role of pragmatics in reflexive interpretation in Korean learners of English. In M. Bowles, R. Foote, S. Perpiñán & R. Bhatt (eds.), *Selected proceedings of the 2007 Second Language Research Forum*, pp. 97–112. Somerville, MA: Cascadilla Proceedings Project.

Marinis, T., Roberts, L., Felser, C. & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27, 53–78.

Matsumura, M. (1994). Japanese learners' acquisition of the locality requirement of English reflexives: Evidence for retreat from overgeneralization. *Studies in Second Language Acquisition*, 16, 19–42.

McElree, B. & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation.*Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 134–157.

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.

Nagata, H. (1995). On-line and off-line reflexive resolution in Japanese logophoric sentences. *Journal of Psycholinguistic Research*, 24, 205–229.

Nicol, J. & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5–20.

Nicol, J. & Swinney, D. (2003). The psycholinguistics of anaphora. In A. Barss (ed.), *Anaphora: A reference guide*, pp. 72–104. Oxford: Blackwell.

Oshima, D. (2007). On empathic and logophoric binding. *Research on Language & Computation*, 5, 19–35.

Pollack, I. & Norman, D. (1964). A non-parametric analysis of signal detection experiments. *Psychonomic Science*, 1, 125–126.

Pollard, C. & Sag, I. A. (1992). Anaphors in English and the scope of binding theory. *Linguistic Inquiry*, 23, 261–303.

Reinhart, T. (1981). Definite NP anaphora and c-command. *Linguistic Inquiry*, 12, 605–635.

Reinhart, T. & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, 24, 657–720.

Roberts, L., Gullberg, M. & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press

effects. *Studies in Second Language Acquisition*, 30, 333–357.

Runner, J., Sussman, R. & Tanenhaus, M. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: Evidence from eye-movements. *Cognition*, 89, B1–B13.

Runner, J., Sussman, R. & Tanenhaus, M. (2006). Assigning reference to reflexives and pronouns in picture noun phrases. *Cognitive Science*, 30, 1–49.

Schütze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

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.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542–562.

Thomas, M. (1991). Universal grammar and the interpretation of reflexives in a second language. *Language*, 67, 211–239.

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

Walther, C. (1995). Processing reflexives in coordinate NPs: A question of point-of-view. *Journal of Psycholinguistic Research*, 24, 39–78.

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.

White, L., Bruhn-Garavito, J., Kawasaki, T., Pater, J. & Prévost, P. (1997). The researcher gave the subject a test about himself: Problems of ambiguity and preference in the investigation of reflexive binding. *Language Learning*, 47, 145–172.

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.

Xiang, M., Dillon, B. & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, 108, 40–55.

Received: February 13, 2008
Final revision received: October 10, 2008
Accepted: December 6, 2008

https://doi.org/10.1017/S1366728909990228 Published online by Cambridge University Press