THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Processing reflexives in a second language: The timing of structural and discourse-level constraints

Citation for published version:
Felser, C & Cunnings, I 2011, 'Processing reflexives in a second language: The timing of structural and discourse-level constraints', *Applied Psycholinguistics*, pp. 1-33.
https://doi.org/10.1017/S0142716411000488

Digital Object Identifier (DOI):
10.1017/S0142716411000488

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Applied Psycholinguistics

Publisher Rights Statement:
©Felser, C., & Cunnings, I. (2011). Processing reflexives in a second language: The timing of structural and discourse-level constraints. Applied Psycholinguistics, 1-33doi: 10.1017/S0142716411000488

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 25. Aug. 2022

Applied Psycholinguistics **33** (2012), 571–603
doi:10.1017/S0142716411000488

Processing reflexives in a second language: The timing of structural and discourse-level constraints

CLAUDIA FELSER and IAN CUNNINGS
University of Essex

Received: October 29, 2009 Accepted for publication: June 20, 2010

ADDRESS FOR CORRESPONDENCE
Claudia Felser, Department of Language and Linguistics, University of Essex, Colchester CO4 3SQ, United Kingdom. E-mail: felsec@essex.ac.uk

ABSTRACT
We report the results from two eye-movement monitoring experiments examining the processing of reflexive pronouns by proficient German-speaking learners of second language (L2) English. Our results show that the nonnative speakers initially tried to link English argument reflexives to a discourse-prominent but structurally inaccessible antecedent, thereby violating binding condition A. Our native speaker controls, in contrast, showed evidence of applying condition A immediately during processing. Together, our findings show that L2 learners' initial focusing on a structurally inaccessible antecedent cannot be due to first language influence and is also independent of whether the inaccessible antecedent c-commands the reflexive. This suggests that unlike native speakers, nonnative speakers of English initially attempt to interpret reflexives through discourse-based coreference assignment rather than syntactic binding.

Previous research on nonnative speakers' ability to apply structural coreference constraints on anaphor resolution has primarily been informed by data from untimed or "offline" tasks. Little is known, in contrast, about the mental processes that underlie second language (L2) learners' interpretations of anaphoric expressions, or the types of information that guide real-time anaphor resolution in nonnative compared to native language processing.[1] Among the factors that may influence the interpretation of pronominal anaphors, including reflexives such as *himself, herself*, and so forth, are gender and number congruence between the anaphor and its antecedent, a potential antecedent's relative discourse prominence, semantic and pragmatic constraints (e.g., plausibility, world knowledge), and phrase structure-based coreference constraints (Nicol & Swinney, 2003).

In the generative grammar tradition, structural constraints on the interpretation of reflexives and pronouns are captured by the principles of binding theory (Chomsky, 1981). Regarding the binding properties of argument reflexives such as *herself* in (1) below, English is more restrictive than many other languages in that

© Cambridge University Press 2011 0142-7164/11 $15.00

the reflexive must normally be bound by the closest c-commanding antecedent (i.e., by *Lisa*).

(1) Jane$_i$ noticed that Lisa$_k$ had hurt herself$_{*i/k}$.

The term *c-command* refers to a relationship between constituents in a phrase structure representation that is based on the hierarchical notion of dominance. According to the standard definition, a constituent c-commands its sister constituents and any constituents that these dominate (Reinhart, 1981). C-command relationships are independent of the linear distance between a potential antecedent and the reflexive, and it is the c-command requirement on reflexive binding that renders *Lisa* the only legitimate binder of *herself* in sentences such as (2), for instance.

(2) Lisa$_i$, who was Jane's$_k$ oldest friend, recently hurt herself$_{i/*k}$ in a road accident.

C-command often coincides with subjecthood, a factor that is known to affect reflexive anaphor resolution in many languages (compare, e.g., Büring, 2005) and that may also render a potential antecedent more discourse prominent.

The results from previous processing studies suggest that native first language (L1) speakers of English apply binding condition A immediately during comprehension (e.g., Nicol & Swinney, 1989; Sturt, 2003). In contrast, there is evidence that L2 learners may violate condition A during early processing stages even if they show nativelike mastery of reflexive binding in offline tasks (Felser, Sato, & Bertenshaw, 2009). Felser et al. (2009) showed that proficient Japanese-speaking learners of English had difficulty applying the locality requirement on English reflexives in processing tasks and initially tried to link reflexive object pronouns to a c-commanding nonlocal antecedent, such as *Jane* in (1) above, during L2 reading. Although influence from the learners' L1 cannot be ruled out altogether here, because Japanese is a language that permits "long-distance" (LD) binding of reflexives by nonlocal antecedents, Felser et al. (2009) speculated that the matrix subject's relative discourse prominence might have been responsible for the learners' nonnativelike processing patterns instead.

Investigating the availability and interaction of structural and nonstructural constraints during L2 anaphor resolution will not only help fill an empirical research gap, but should furthermore help us evaluate the claim that nonnative comprehenders' ability to use structural cues to interpretation during processing may be limited, compared to their ability to make use of nonstructural cues (Clahsen & Felser, 2006). According to Clahsen and Felser's (2006) "shallow structure hypothesis," L2 learners have difficulty building complex structural representations of the L2 input in real time, but may be able to compensate for this by making efficient use of nonstructural information instead. Alternatively, L1/L2 processing differences have been attributed to general cognitive factors such as slower L2 processing speed or the increased computational resource demands associated with processing a nonnative language (e.g., McDonald, 2006) or to a lack of proficiency and/or L1 influence (e.g., Hopp, 2006). Building on and extending Felser et al.'s (2009) work, the current study seeks to determine more

precisely what information sources guide learners' initial antecedent preferences when a reflexive is first encountered in the input, while eliminating L1 influence and slower L2 processing speed as potentially confounding variables.

CONDITION A IN NATIVE LANGUAGE PROCESSING

Native English speakers' adherence to binding condition A during online anaphor resolution is rather well attested, at least for argument reflexives of the type under investigation here (e.g., Harris, Wexler, & Holcomb, 2000; Nicol & Swinney, 1989; Sturt, 2003; Xiang, Dillon, & Phillips, 2009).[2] In recent L1 processing studies the binding-theoretically appropriate antecedent is usually referred to as the "accessible" antecedent, whereas a structurally inappropriate competitor antecedent is referred to as the "inaccessible" antecedent, a labeling convention that we will adopt in the remainder of this article.

According to the "binding as initial filter" hypothesis, condition A immediately rules out any structurally inaccessible antecedents from the candidate set. Evidence for this comes from a cross-modal priming study by Nicol and Swinney (1989), who found that upon encountering a reflexive in sentences such as *The boxer told the skier that the doctor for the team would blame himself for the recent injury*, participants would mentally reactivate the accessible antecedent (i.e., *the doctor*) only. The binding as initial filter hypothesis has subsequently been qualified by Sturt (2003) on the basis of the results from an eye-movement monitoring study. This technique provides a detailed reading-time record that includes both "early" eye-movement measures such as first fixations or first-pass reading times, and "later" measures such as the time participants spend reading a given sentence region again after their eyes had already moved away from it. Early measures are thought to provide information about initial analysis and information integration processes, whereas second-pass or rereading times are sensitive to later processes such as reanalysis or discourse integration (Frenck-Mestre, 2005; Staub & Rayner, 2007).

Using materials such as those shown in (3), with gender-stereotype violations used as diagnostics for coreference assignment, Sturt's results suggest that English speakers may violate condition A under certain conditions, at later processing stages.

(3) a. Jennifer was pretty worried at the City Hospital. She remembered that the surgeon had pricked *herself* with a used syringe needle.
b. Jennifer was pretty worried at the City Hospital. She remembered that the surgeon had pricked *himself* with a used syringe needle.

In both (3a) and (3b), the reflexive is syntactically bound by the local noun phrase *the surgeon*. In (3a), the reflexive *herself* mismatches the stereotypical gender of the occupational noun *surgeon*, which is typically taken to be male, whereas in (3b) there is a gender mismatch between the masculine reflexive *himself* and the inaccessible antecedent *Jennifer.* Sturt found evidence for the early application of binding condition A during processing in the shape of shorter initial fixations on the reflexive when the reflexive matched the accessible antecedent's stereotypical gender (as in (3b)) compared to when it did not. Together with the absence of

any effects of the inaccessible antecedent's gender (*Jonathan/he* vs. *Jennifer/she*) during participants' first reading of the reflexive, this indicates that readers initially tried to link the reflexive to its local antecedent while disregarding the inaccessible one.

Some effects of the inaccessible antecedent were observed at later processing stages, though, as witnessed by interactions between the factors "accessible" and "inaccessible" antecedent in participants' second-pass reading times at the reflexive and prefinal sentence regions. Specifically, gender-mismatching inaccessible antecedents led to elevated reading times compared to matching inaccessible ones in the two "accessible-match" conditions. Together, these findings led Sturt to propose that binding condition A, while constraining the antecedent search initially, might be defeasible during later processing stages.

The results from a follow-up experiment confirmed that participants' initial focusing on the accessible antecedent was not simply due to the accessible antecedent's linear proximity to the reflexive. Using slightly modified materials in which the inaccessible antecedent was linearly closer to, but failed to c-command, the reflexive (e.g., *The surgeon who treated Jennifer had pricked himself/herself. . . .*), Sturt (2003) replicated the early accessible antecedent-mismatch effect (for *the surgeon . . . herself*) but did not find any statistically reliable effects of the inaccessible antecedent (*Jennifer*). Taken together, Sturt took the results from his two experiments to suggest that only highly discourse-prominent competitor antecedents might be considered besides the structurally accessible one during later processing stages.[3]

Possible effects of a linearly closer but structurally inaccessible competitor antecedent were also examined by Xiang et al. (2009) using event-related potentials (ERPs). Their design included three experimental conditions as exemplified by (4a–c), with materials that differed from those used in Sturt's (2003) second experiment in that both potential antecedents were structural subjects. Only *the well-known surgeon* is an accessible antecedent for the reflexive here, as the linearly closer competitor antecedent (*Jonathan/Jennifer*) does not c-command the reflexive.

(4) a. The well-known surgeon that Jonathan had studied with in school pricked himself with a used needle.
b. *The well-known surgeon* that Jennifer had studied with in school pricked *herself* with a used needle.
c. *The well-known surgeon* that *Jonathan* had studied with in school pricked *herself* with a used needle.

Besides a "double-match" baseline condition (4a), Xiang et al.'s (2009) experimental sentences contained stereotypical gender mismatches either between the reflexive and the accessible antecedent (4b), or between the reflexive and both potential antecedents (4c). Both mismatch conditions (4b) and (4c) elicited a "P600," a brain response thought to index syntactic processing or information integration difficulty, compared to the "double match" condition (4a), when participants read the reflexive. There was no reliable difference between the "accessible mismatch" condition (4b) and the "double mismatch" condition (4c), suggesting that only the structurally accessible antecedent was considered. This is consistent

with the results from Sturt's (2003) second experiment and suggests that non-c-commanding (or less discourse-prominent) inaccessible antecedents may not be considered at all during L1 processing.

Summarizing, the results from L1 processing studies using time-course sensitive experimental techniques such as eye-movement monitoring or ERPs have shown that binding condition A constrains the initial search for an antecedent for argument reflexives. There is some evidence that a highly salient but inaccessible antecedent may also be considered, albeit only at later processing stages.

ONLINE ANAPHOR RESOLUTION IN NONNATIVE LANGUAGE PROCESSING

As most previous L2 acquisition studies have focused on learners' sensitivity to structural constraints such as the locality requirement on reflexive binding in untimed tasks (for a review and discussion, see Hawkins, 2001; White, 2003), the role of nonstructural factors in L2 anaphor resolution has thus far received little attention. There is some evidence that learners' offline interpretation of reflexives may be more strongly affected by pragmatic or discourse information than native speakers' (Demirci, 2000, 2001; Lee, 2008), but the extent to which structural and discourse-level information affect online reference resolution in an L2 has only recently begun to be investigated (Felser et al., 2009; Roberts, Gullberg, & Indefrey, 2008).

Using a timed ("paced") grammaticality judgment task, Felser et al. (2009) found that proficient Japanese-speaking learners of L2 English were significantly less accurate than native speakers at detecting accessible antecedent mismatches in sentences like (6a), which contained a matching but inaccessible competitor antecedent (i.e., *Mary*) in matrix subject position. No reliable L1/L2 differences were observed in participants' ability to judge sentences of the type shown in (5b), where the matching inaccessible antecedent did not c-command the reflexive.

(5) a. *Mary believed that *the dancers* had hurt *herself*.
b. *The dancers believed that Mary's *brother* had hurt *herself*.

Unlike the native controls, the learners also took significantly longer to judge locality violations, as in (5a), compared to c-command violations, as in (5b). These findings suggest that unlike the native controls, the L2 learners were more likely to take an inaccessible antecedent into account if it c-commanded the reflexive, as in (5a), than when it did not, as in (5b).

This was corroborated by the results from a second experiment using eye-movement monitoring during reading. Materials included brief paragraphs consisting of a lead-in sentence that introduced two named referents (e.g., *John* and *Richard*), followed by a critical sentence containing a reflexive and a final "closing" sentence, as shown in (6).

(6) John/Jane and Richard were very worried in the kitchen of the expensive restaurant.
a. *John/Jane* noticed that Richard had cut *himself* with a very sharp knife.
b. It was clear to *John/Jane* that Richard had cut *himself* with a very sharp knife.
Kitchens can be dangerous places.

The inaccessible antecedent's gender was manipulated so that it either matched (*John*) or mismatched (*Jane*) the reflexive's morphological gender. In addition, Felser et al. (2009) manipulated the critical sentences' syntactic form so that the inaccessible antecedent either c-commanded (6a) or failed to c-command (6b) the reflexive. Note that in sentences of type (6b), the inaccessible antecedent's relative discourse-salience is also reduced compared to the former.

The analysis of the reading time data revealed that the L2 group but not the native speakers were affected by the inaccessible antecedent's gender during their initial reading of the reflexive, an effect that was restricted to those conditions in which the inaccessible antecedent was in matrix subject position, as in (6a). The learners showed significantly longer first-pass reading times when the matrix subject matched the reflexive in gender compared to when it did not, which indicates that they were confused by the presence of two matching c-commanding antecedents. Their ultimate interpretations of English reflexives as measured by an offline antecedent choice task, in contrast, were nativelike and in line with condition A.

Unlike in Sturt's (2003) Experiment 1, the English native speakers in Felser et al.'s study did not show any effects of the inaccessible antecedent's gender in later eye-movement measures either. Whereas the absence of any effects of the inaccessible antecedent in the native group's reading times is consistent with the early application of condition A in L1 processing, the results from the nonnative speakers suggest that they considered a structurally and discourse-prominent but inaccessible antecedent during early processing stages. However, as Felser et al. (2009) only manipulated the inaccessible, but not the accessible, antecedent in their study, their results do not tell us anything about whether or when during processing the accessible antecedent was considered. Thus, the relative timing of effects of the accessible and inaccessible antecedents in L2 processing is an issue still in need of further investigation.

Given that the L2 groups in Felser et al.'s (2009) study consisted of learners whose native language (Japanese) permits LD binding of reflexives, it is tempting to account for their temporary confusion during the processing of English reflexives in terms of L1 influence. Although Felser et al. (2009, pp. 498–499) present some arguments against explanations in terms of (morpho-)syntactic L1 transfer, it is conceivable that some form of L1 influence might have occurred at the discourse-pragmatic level. That is, the learners' initial attempt at reference resolution might have been affected by the inaccessible antecedent's subject- or topichood, with L1 discourse-based preferences interacting with the locality requirement on English reflexive binding. The role of L1 influence in nonnative speakers' processing of reflexives thus clearly requires more investigation.

Alternatively, Felser et al.'s (2009) findings might reflect general L2 processing effects, such as a stronger reliance on discourse-level information in L2 compared to L1 processing. Evidence that discourse-pragmatic information affects online anaphor resolution in the L2 independently of learners' L1 has been reported by Roberts et al. (2008), who examined Turkish- and German-speaking learners' offline and online interpretation of pronouns in L2 Dutch, using materials such as those shown in (7) below.

(7) a. De werkneemers zitten in het kantoor. Terwijl Peter aan het werk is, eet hij een boterham.
"The workers are in the office. While Peter is working, he is eating a sandwich."
b. Peter en Hans zitten in het kantoor. Terwijl Peter aan het werk is, eet hij een boterham.
"Peter and Hans are in the office. While Peter is working, he is eating a sandwich."

Pronouns differ from argument reflexives in that they need not be syntactically bound but can also be linked to a potential referent within or outside the current sentence via discourse-based coreference assignment (e.g., Reinhart, 1983). In (7a), the matrix subject *hij* "he" is most likely to be resolved locally and identified with *Peter*, the subject of the preceding adjunct clause. The pronoun cannot be interpreted as being coreferential with *de werkneemers* "the workers" in the first sentence because of the number mismatch. The initial sentence in (7b), in contrast, introduces two potential (i.e., gender- and number-matched) referents for *hij*, *Peter* and *Hans*. The two L2 groups' reading-time patterns were similar, and differently from the native Dutch speakers', in showing longer second-pass and total reading times when the context provided two alternative referents, as in (7b), compared to when it did not. This appears to suggest that the learners were confused, during their rereading of the pronoun region, when a matching competitor antecedent was available in the discourse context.

Despite differences in timing, in that the effects of a matching competitor antecedent provided in the discourse context were seen in early eye movement measures in Felser et al.'s (2009) study but were restricted to later measures in Roberts et al.'s (2008) study, the above findings suggest that online reference resolution may be more strongly guided by discourse information in nonnative than in native language processing. Exactly when during processing discourse-level information affects L2 anaphor resolution, and how this interacts with structural coreference constraints, is still rather unclear, however. Using a time course-sensitive experimental method such as eye-movement monitoring will allow us to investigate and compare the relative timing of structural and nonstructural constraints during native and nonnative comprehension.

EXPERIMENT 1

Our first experiment investigates whether learners' taking into account a nonlocal competitor antecedent for a reflexive is contingent on the availability of LD binding in their native language, by testing learners whose L1 is similar to English in that argument reflexives must also be bound locally. By manipulating gender congruence both with the accessible and the inaccessible antecedent, the presence and timing of accessible or inaccessible gender effects should allow us to determine not only whether, but also when, during processing each of the two potential antecedents are considered.

Method

Participants. Participants were 25 adult German-speaking learners of L2 English (7 males, mean age = 24.4) and 28 native English-speaking controls (14 males,

mean age = 22.4), recruited from the University of Essex student and staff communities. All had normal or corrected to normal vision. They received a small fee for their participation and were not informed of the ultimate purpose of the main experiment. The nonnative participants had all started learning English at school between ages 8 and 13 and had spent 32.68 months on average (SD = 49.50 months) in the United Kingdom at the time of testing. To obtain a measure of the learners' general level of English proficiency at the time of testing, they were asked to complete the computerized version of the Quick Placement Test (Oxford University Press, 2001). The learner group's mean test score was 82.64% (SD = 11.87%), with the majority (n = 17) falling within the "upper advanced" bracket and the remaining ones classed as "lower intermediate" (n = 1), "upper intermediate" (n = 3), or "lower advanced" (n = 4) learners.

To assess their knowledge of binding condition A, all participants underwent an offline multiple-choice antecedent identification task, which was administered after the main experiment in the form of a written questionnaire. Materials included 12 sets of sentences containing reflexive and a further 12 containing nonreflexive pronouns. The inaccessible antecedent's gender was manipulated so that it either matched or mismatched the pronoun's gender, as illustrated in (8a–d) below.

(8) a. REFLEXIVE, SINGLE MATCH
Emma noticed that the grandfather had explained himself carefully.
b. REFLEXIVE, DOUBLE MATCH
Adam noticed that the grandfather had explained himself carefully.
c. PRONOUN, SINGLE MATCH
Daniel recalled that Emma had woken him too late.
d. PRONOUN, DOUBLE MATCH
Daniel recalled that Adam had woken him too late.

The linear order of the two potential antecedents in the reflexive conditions matched the linear order of the potential antecedents in the online experiment, which are described further below. The 24 test items were pseudorandomized with 12 ambiguous fillers (e.g., *Susan told Mary a story about herself*), and participants had to indicate which of the potential referents mentioned they thought the reflexives or pronouns referred to. For each item, participants were offered three answer choices: the accessible antecedent (e.g., *the grandfather* in (8a,b)), the inaccessible antecedent (*Emma/Adam*), or either of them. Participants hardly ever provided any "either" responses. Across all conditions, the native speakers gave "either" responses 1.3% of the time and the German learners 0.5% of the time. Table 1 provides a summary of participants' correct answer choices across the four conditions. Both the native and the nonnative participants performed at or near ceiling in this task, confirming that they knew the binding properties of English argument reflexives and were aware of the different binding requirements of reflexives and pronouns.

Materials. The materials used for the online task included 24 sets of items in four conditions, modeled after Sturt (2003; Experiment 1). Each experimental

Table 1. *Native speakers (NSs) and nonnative speakers' (NNSs) mean (standard deviation) percentages of correct choices of the accessible antecedent per condition in the offline antecedent identification task in Experiment 1*

	Reflexive	Pronoun
NSs ($n = 28$)		
Single match	100 (0)	98 (7)
Double match	98 (7)	92 (15)
NNSs ($n = 25$)		
Single match	100 (0)	98 (7)
Double match	99 (3)	95 (14)

item consisted of a lead-in sentence that introduced a named referent and served to "set the scene," a critical sentence containing a reflexive, and a wrap-up sentence. Following Sturt (2003), we manipulated both the stereotypical gender congruence between the accessible antecedent and the reflexive, and the gender congruence between the inaccessible antecedent and the reflexive, as illustrated by (9a–d).[4]

(9) a. ACCESSIBLE MATCH, INACCESSIBLE MATCH
James has worked at the army hospital for years. He noticed that the soldier had wounded himself while on duty in the Far East. Life must be difficult when you are in the army.

b. ACCESSIBLE MATCH, INACCESSIBLE MISMATCH
Helen has worked at the army hospital for years. She noticed that the soldier had wounded himself while on duty in the Far East. Life must be difficult when you are in the army.

c. ACCESSIBLE MISMATCH, INACCESSIBLE MATCH
Helen has worked at the army hospital for years. She noticed that the soldier had wounded herself while on duty in the Far East. Life must be difficult when you are in the army.

d. ACCESSIBLE MISMATCH, INACCESSIBLE MISMATCH
James has worked at the army hospital for years. He noticed that the soldier had wounded herself while on duty in the Far East. Life must be difficult when you are in the army.

The proper names were all common English male or female names, with each pair (e.g., *James/Helen*) matched for length. The reliability of the stereotype manipulation was tested in a gender-rating questionnaire that all participants completed after the main eye-movement experiment. Participants rated the gender of the 12 stereotypically male and 12 stereotypically female nouns used in the main experiment, plus 12 additional "filler" nouns that were gender neutral (e.g., *adult*). Participants rated each noun on a scale from 1 (*highly likely to refer to a female*)

to 7 (*highly likely to refer to a male*). For vocabulary screening purposes, "Do not know this word" was also included as an option. Both the native speakers and L2 learners rated the female stereotyped nouns (which received average scores of 2.57 and 2.85, respectively) lower on the scale than the male stereotyped nouns (which received scores of 5.80 and 5.51, respectively). The *t* tests revealed this difference to be reliable for both groups: for native speakers, $t_1\ (27) = 15.58$, $p < .001$; $t_2\ (21) = 18.89$, $p < .001$; for L2 learners, $t_1\ (24) = 10.48$, $p < .001$; $t_2\ (21) = 15.84$, $p < .001$.

The experimental items were distributed across four presentation lists in a Latin-square design, mixed with 56 filler texts and pseudorandomized. Among the fillers were 12 distracter items containing different types of anaphor (pronouns or reflexives) in structurally different positions to those in the experimental stimuli. 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 questions following the critical items never directly probed into participants' interpretation of the reflexive.

The predictions for this experiment are as follows. Given previous findings by Sturt (2003) and others, we expect native speakers' initial processing of the reflexive to be constrained by binding condition A. This should be reflected in main effects of the accessible antecedent (i.e., *the soldier*) in early eye-movement measures, with longer reading times in the accessible mismatch (9c,d) compared to the "accessible match" conditions (9a,b). Effects of the inaccessible antecedent (*James . . . he* vs. *Helen . . . she*) should be either absent or restricted to later processing measures or sentence regions.

Regarding nonnative speakers, if the early effect of the inaccessible antecedent's gender observed in Felser et al.'s (2009) nonnative participants merely reflected some form of transfer of LD binding from L1 Japanese to L2 English, then main effects of the inaccessible antecedent should be absent from our German group's early eye-movement measures. Instead, German learners of English should pattern with the native controls in showing early sensitivity to Condition A, in the form of a main effect of the accessible antecedent's gender. Note that the German translation equivalents of (9a–d) also require local binding of the reflexive, thus precluding the possibility of L1 transfer of LD binding.

(10)

Er bemerkte,	*dass*	*[der Soldat]$_i$*	*sich$_i$*	*während*	*des*	*Dienstes*	*im*
He noticed	that	the soldier	REFL	during	the	duty	in.the
Fernen	*Osten*	*verwundet*	*hatte.*				
Far	East	wounded	had				

"He noticed that the soldier had wounded himself while on duty in the Far East."

In (10), for instance, the reflexive pronoun *sich* can only be linked to the local subject *der Soldat* "the soldier" but not to the matrix subject pronoun *er* "he." The German reflexive *sich* (as well as its emphatic form, *sich selbst* "REFL self") differs from English reflexives in being unmarked for gender. The learners' nativelike performance in the offline questionaire task confirmed that they were sensitive to the gender properties of English reflexives, however.

If, in contrast, the initial antecedent search in nonnative anaphor resolution is not guided by syntactic coreference constraints but instead relies primarily on other cues to interpretation such as discourse prominence, then main effects of the inaccessible antecedent's gender (9a,c < 9b,d) might in fact be seen earlier than main effects of the accessible antecedent (9a,b < 9c,d) in the German group, or the two factors might interact in early eye-movement measures (as was seen in some of the later measures in Sturt, 2003).

Procedure. All participants were tested individually in a quiet, dedicated laboratory room. Their eye movements were recorded using a head-mounted SR Research Eyelink II system, with a sample resolution of 500 Hz. Participants' eye movements were monitored during their reading of the experimental texts on a computer screen by two miniature cameras mounted on a headband. All items were presented in Courier New font in black letters on white background and displayed across three lines of text.

Although participants read binocularly, we recorded only information from the right eye. The system automatically compensated for head movements through the use of another camera at the center of the headband. Each experimental session started with a calibration procedure, and calibration was checked again before each trial. Any possible drift in the headset was automatically compensated for before a new stimulus paragraph was presented.

The stimulus presentation was divided into four blocks, allowing participants to take up to three breaks. Forward and reverse orders within each block were constructed, with each order being completed by half of the subjects, and the ordering of each block was different for each participant. The experiment began with five practice items to familiarize participants with the procedure. Participants were instructed to read each stimulus text silently for comprehension at a comfortable reading speed, and to indicate by a button press when they had finished. The end of trial comprehension questions required a binary yes/no push-button response.

The native participants completed the experiment within approximately 30–40 min, and the learners were tested in two separate sessions of similar duration. The eye-movement experiment, antecedent choice task, and gender-rating questionnaire were carried out during the first session, and the Quick Placement Test in the second.

Data trimming and analysis. Results will be reported for four eye movement measures, first fixation durations, first-pass reading times, regression path durations, and rereading times, for both the critical reflexive region (*himself/herself*) and the postcritical region, which included the two words following the reflexive (e.g., *while on*). First fixation duration is the length of readers' initial fixation on a given interest region, and first-pass reading time is the summed duration of all fixations on a region until that region is first exited to the left or right. Regression path duration is the sum of all fixations on a region until the eyes move on to the right, and thus may also include regressive eye movements to earlier sentence regions. Rereading time is the summed duration of all fixations within a region after this region was exited to either the left or right. Prior to the analysis of the reading time data, fixations shorter than 80 ms were merged with a neighboring fixation if this

Table 2. *Native speakers (NSs) and nonnative speakers' (NNSs) mean (standard deviation) reading times in the reflexive region (ms) in Experiment 1*

	1st Fixation Durations	1st Pass Reading Time	Regression Path Duration	Rereading Time
Accessible match, inaccessible match (=9a)				
NSs	226 (42)	241 (84)	309 (130)	150 (117)
NNSs	226 (30)	244 (33)	298 (89)	160 (139)
Accessible match, inaccessible mismatch (=9b)				
NSs	216 (43)	243 (95)	313 (124)	200 (114)
NNSs	238 (35)	253 (47)	298 (84)	114 (102)
Accessible mismatch, inaccessible match (=9c)				
NSs	234 (47)	252 (61)	341 (157)	223 (111)
NNSs	224 (33)	239 (39)	285 (73)	177 (141)
Accessible mismatch, inaccessible mismatch (=9d)				
NSs	230 (34)	254 (77)	357 (221)	235 (150)
NNSs	242 (30)	265 (43)	333 (100)	192 (166)

was within one degree of another fixation. All other fixations below 80 ms, and all those above 800 ms, were deleted before any further analysis. Trials in which track loss occurred, or when a region of text was initially skipped, were treated as missing data. For each reading time measure at each region, outliers 2.5 *SD* above or below the participant mean for that measure at that region were removed.

To establish whether the two participant groups' reading times across the four experimental condition were statistically different, we initially carried out a series of preliminary mixed analyses of variance (ANOVAs) with the within-subjects factors accessible antecedent (*match, mismatch*) and inaccessible antecedent (*match, mismatch*), and the between-subjects factor group (*L1, L2*). For sentence regions where main effects of, or interactions with, the factor group were observed, we went on to analyze the data from each participant group separately.

Results

Participants' overall comprehension accuracy was very good, with the native group answering 91.76% and the German group 95.15% of the comprehension questions correctly. This confirms that both groups were actively reading the experimental paragraphs for meaning.

There was no track loss in either the English or German data, but 0.17% of trials for the German group were removed as participants reported not knowing the critical vocabulary items (the gender stereotyped nouns) in the gender rating questionnaire. Skipping rates for the reflexive region were 11.2% for the native group and 1.67% for the learners, and the postcritical region was skipped 5.06% of the time by the native speakers and 0.67% by the learners. Outlier removal resulted in the loss of no more than 4.5% of the data for each measure and region. Table 2 and Table 3 summarize participants' reading times across the four experimental conditions.

Table 3. *Native speakers (NSs) and nonnative speakers' (NNSs) mean (standard deviation) reading times in the postcritical region (ms) in Experiment 1*

	1st Fixation Durations	1st Pass Reading Time	Regression Path Duration	Rereading Time
Accessible match, inaccessible match (=9a)				
NSs	241 (59)	326 (144)	428 (194)	144 (105)
NNSs	223 (40)	359 (76)	396 (109)	178 (179)
Accessible match, inaccessible mismatch (=9b)				
NSs	238 (36)	316 (91)	479 (161)	201 (135)
NNSs	228 (27)	387 (91)	426 (121)	155 (141)
Accessible mismatch, inaccessible match (=9c)				
NSs	235 (31)	328 (91)	539 (255)	225 (129)
NNSs	235 (32)	368 (86)	474 (169)	232 (184)
Accessible mismatch, inaccessible mismatch (=9d)				
NSs	235 (48)	317 (125)	573 (238)	238 (153)
NNSs	238 (32)	379 (102)	484 (159)	192 (215)

To see whether there were any statistical L1/L2 differences in participants' reading times, we carried out preliminary mixed ANOVAs as described above, for both regions of interest. At the reflexive region, these showed interactions with the factor group, which were significant in the analysis by participants for first fixation durations: Inaccessible × Group, $F_1\ (1, 52) = 12.14, p < .01$; $F_2\ (1, 23) = 2.91, p = .101$, marginally significant by participants for first-pass reading times: Inaccessible × Group, $F_1\ (1, 52) = 3.70, p = .060$; $F_2\ (1, 23) = .60, p = .447$, and significant by both participants and items for rereading times: Accessible × Inaccessible × Group, $F_1\ (1, 52) = 4.72, p < .05$; $F_2\ (1, 23) = 4.67, p < .05$. Main effects of group were found only in regression path durations, $F_1\ (1, 52) = .88, p = .354$; $F_2\ (1, 23) = 4.80, p < .05$, and rereading times, $F_1\ (1, 52) = 2.27, p = .138$; $F_2\ (1, 23) = 9.97, p < .01$, in the analyses by items, reflecting that the nonnative speakers tended to read the reflexive region slightly faster than the native controls overall.

At the postcritical region, the factor group was again found to modulate participants' reading times in first fixation durations, in the analysis by participants: Accessible × Group, $F_1\ (1, 52) = 7.05, p < .05$; $F_2\ (1, 23) = 2.94, p = .100$, first-pass reading times, Inaccessible × Group, $F_1\ (1, 52) = 4.13, p < .05$; $F_2\ (1, 23) = 4.90, p < .05$, and rereading times, marginally so by items, Inaccessible × Group, $F_1\ (1, 52) = 4.64, p < .05$; $F_2\ (1, 23) = 3.18, p = .088$. Significant main effects of group were seen in first-pass, $F_1\ (1, 52) = 4.04, p = .050$; $F_2\ (1, 23) = 27.35, p < .001$, and regression path durations, $F_1\ (1, 52) = 2.30, p = .136$; $F_2\ (1, 23) = 14.75, p < .01$, in the analyses by items.

As the above effects of group are indicative of differences between the native speakers' and the learners' reading-time patterns across the experimental conditions, we then analyzed each participant group's data separately.

Native speakers. A series of repeated-measures ANOVAs with the factors accessible antecedent (*match, mismatch*) and inaccessible antecedent (*match, mismatch*) were carried out for each eye-movement measure. Table 4 provides a summary of the results.

At the reflexive region, the native group generally showed shorter reading times when the accessible antecedent matched the reflexive in gender (=9a,b) compared to when it did not (=9c,d). This was reflected statistically in significant main effects of the accessible antecedent in first fixation durations in the analysis by participants, in regression path durations in the analysis by items, as well as in the rereading times. There were no other significant main effects or interactions. The same pattern was seen at the postcritical region, with significant main effects of the accessible antecedent seen in regression path durations and rereading times, in the absence of any other significant effects or interactions. The results from the control group thus support our prediction that binding condition A should constrain native speakers' initial interpretation of argument reflexives.

L2 learners. At the reflexive region, the German group patterned differently from the native controls in showing longer reading times when the inaccessible antecedent mismatched the reflexive's gender compared to when it did not, in early eye-movement measures, including their first fixation durations (compare Figure 1). Table 5 summarizes the L2 group's ANOVA results.

Significant main effects of the inaccessible antecedent (9a,c $<$ 9b,d) were found in first fixation durations in the analysis by participants, marginal by items, and in first-pass reading times, in the analysis by participants. Effects of the accessible antecedent, in contrast, were not visible until learners' rereading of the reflexive, marginal in the analysis by participants. This effect was modulated by an interaction with the factor inaccessible antecedent in the analysis by items, marginally so in the participants analysis. Subsequent *t* tests showed that condition (9b) (accessible match, inaccessible mismatch) tended to have shorter rereading times than the other conditions: (9b) versus (9a), 114 versus 160 ms; $t_1\ (24) = 1.71$, $p = .099$; $t_2\ (23) = 2.53$, $p < .05$; (9b) versus (9c): 114 versus 177 ms; $t_1\ (24) = 2.30$, $p < .05$; $t_2\ (23) = 2.64$, $p < .05$; (9b) versus (9d): 114 versus 192 ms; $t_1\ (24) = 2.60$, $p < .05$; $t_2\ (23) = 3.31$, $p < .01$, with no other significant differences.

At the postcritical region, significant main effects of the accessible antecedent (9a,b $<$ 9c,d) were found for first fixation durations, marginal by items, regression path durations, and rereading times in the analysis by participants. No other significant main effects or interactions were found at this region. Together, these results indicate that our nonnative participants initially tried to link the reflexive to the inaccessible antecedent, with the accessible one being considered only at later processing stages.

To explore the possibility that the observed L1/L2 processing differences at the critical reflexive region were caused by the less proficient learners, with those at the higher end of the scale showing a more nativelike reading time pattern, we carried out additional analyses with just those learners who fell within the "upper advanced" proficiency category ($n = 17$). The pattern of results was essentially the same as in the main analysis, the only difference being that for the "upper advanced" subgroup, effects of the accessible antecedent started to emerge slightly earlier than in the group as a whole, namely, in the by-participants analysis of the

Table 4. *Summary of analysis of variance results for the reflexive and postcritical regions in Experiment 1 for native speakers*

		Reflexive Region				Postcritical Region			
		1st Fixations	1st Pass Times	Regression Path	Rereading Times	1st Fixations	1st Pass Times	Regression Path	Rereading Times
Accessible	$F1$ (1, 27)	4.63*	2.07	2.78	13.60**	1.15	<1	9.01**	7.30*
Antecedent	$F2$ (1, 23)	1.89	<1	4.77*	23.86**	<1	<1	10.72**	6.43*
Inaccessible	$F1$ (1, 27)	2.30	<1	<1	1.67	<1	1.63	2.71	2.29
Antecedent	$F2$ (1, 23)	<1	<1	<1	3.34†	<1	<1	3.00†	3.16†
Interaction	$F1$ (1, 27)	<1	<1	<1	1.60	<1	<1	<1	1.71
	$F2$ (1, 23)	<1	<1	<1	1.11	<1	<1	<1	<1

†$p < .1$. *$p < .05$. **$p < .01$.

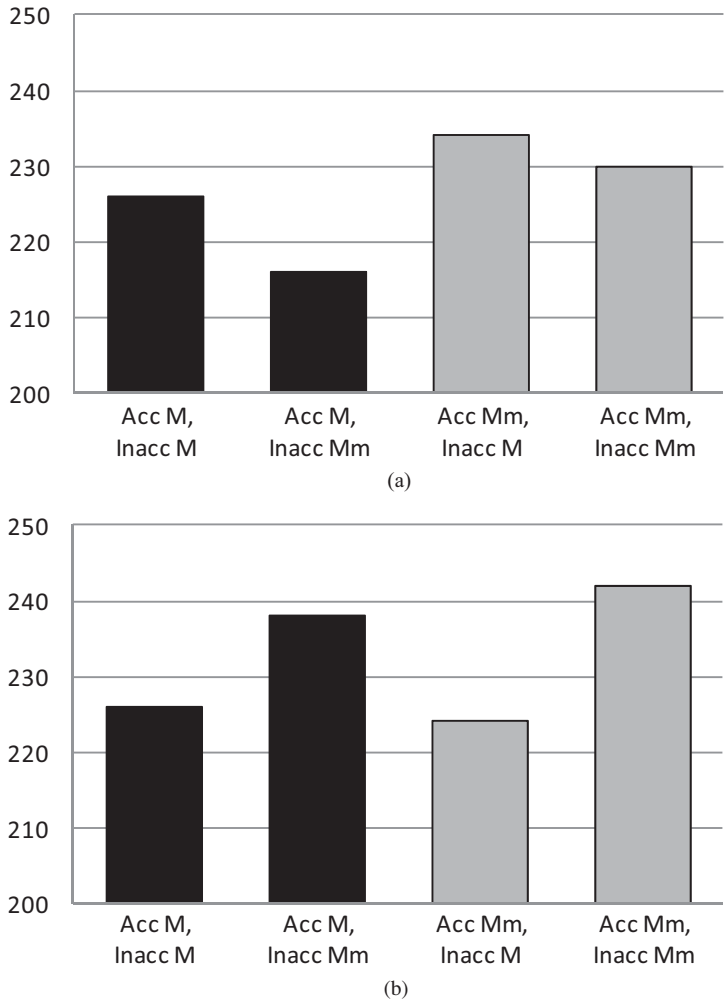

Figure 1. (a) Native and (b) nonnative speakers' mean first fixation durations (ms) on the reflexive in Experiment 1; Acc M, accessible match; Acc Mm, accessible mismatch; Inacc M, inaccessible match; Inacc Mm, inaccessible mismatch.

regression path times, F_1 (1, 16) = 6.13, $p < .05$; F_2 (1, 23) = 2.81, $p = .107$. That is, even the most advanced learners initially considered only the inaccessible antecedent when first encountering a reflexive object pronoun.

Discussion

The results from Experiment 1 show that despite being nativelike in their offline antecedent choices, the German-speaking participants temporarily violated

Table 5. *Summary of analysis of variance results for the reflexive and postcritical regions in Experiment 1 for second language learners*

		Reflexive Region				Postcritical Region			
		1st Fixations	1st Pass Times	Regression Path	Rereading Times	1st Fixations	1st Pass Times	Regression Path	Rereading Times
Accessible	$F1$ (1, 24)	<1	<1	<1	3.80†	8.02**	<1	10.51**	7.75*
Antecedent	$F2$ (1, 23)	<1	<1	<1	8.36**	3.10†	<1	8.24**	2.52
Inaccessible	$F1$ (1, 24)	11.96**	9.83**	2.58	0.80	<1	2.38	1.31	2.50
Antecedent	$F2$ (1, 23)	4.08†	2.75	2.23	1.13	<1	2.36	1.51	1.39
Interaction	$F1$ (1, 24)	<1	1.43	2.13	3.17†	<1	1.28	<1	<1
	$F2$ (1, 23)	<1	1.67	2.84	4.90*	<1	<1	<1	<1

†$p < .1$. *$p < .05$. **$p < .01$.

binding condition A when coming across a reflexive during L2 reading. Unlike the native control group, who showed the expected main effects of the accessible antecedent's gender on the reflexive region in the absence of any effects of the inaccessible antecedent's gender, the learners initially showed effects of the inaccessible antecedent's gender only. Effects of the accessible antecedent were delayed until later processing stages, and were most clearly in evidence during the learners' reading of the postcritical region.

The results from the native speakers corroborate those from earlier L1 processing studies (e.g., Nicol & Swinney, 1989; Sturt, 2003), indicating that condition A is applied immediately in native language comprehension. There was a weak trend toward later effects of the inaccessible antecedent during the native speakers' rereading of the reflexive and during their (re-)reading of the postcritical region, in line with the delayed inaccessible-mismatch effects reported by Sturt, but these were not statistically reliable.

In contrast, our L2 group behaved in a way that has never been observed in native speakers. The results from our German speakers are consistent with those reported by Felser et al. (2009) for Japanese-speaking learners of English, insofar as main effects of the inaccessible antecedent were observed during learners' first reading of the reflexive region. As German argument reflexives also require local binding, the learners' initial preference for a nonlocal antecedent cannot be accounted for by L1 transfer. However, recall that unlike the L2 group examined in Felser et al.'s (2009) second experiment, the German learners in the current experiment showed an inaccessible "mismatch" rather than a "match" effect. That is, although Felser et al.'s (2009) learners were slowed down by the presence of a gender matching inaccessible antecedent, our L2 group was slowed down if the inaccessible antecedent mismatched the reflexive in gender. This could be taken to suggest that the Japanese learners in Felser et al.'s (2009) study were experiencing temporary competition between the two potential antecedents, whereas our German group simply ignored the accessible one during early processing stages.

This in turn could be due to differences between the materials used, with the inaccessible antecedent's relative discourse prominence being higher in the current study than in Felser et al.'s (2009). Recall that both potential antecedents were introduced together in the lead-in sentence in Felser et al.'s (2009) study, whereas only the inaccessible one was initially introduced into the discourse in our experimental materials.

That our German-speaking participants first considered a nonlocal rather than the local antecedent suggests that an antecedent's relative linear proximity to the reflexive did not affect their initial anaphor resolution attempts. It is conceivable, however, that our learners initially preferred to link the reflexive to a named referent rather than to a definite description, in violation of condition A, due to the gender properties of proper names being easier for them to resolve online than those of occupational nouns. However, if names enjoyed a general processing advantage over full noun phrases, then it is unclear why the native controls should have shown immediate sensitivity to our stereotypical gender mismatches but not to name–pronoun mismatches (compare also Sturt, 2003). Note also that for almost all of the occupational nouns that we used, the grammatical gender of the German translation equivalents matched their gender stereotype (e.g., *the*

soldier – der_{masc} *Soldat*, *the midwife* – die_{fem} *Hebamme*).[5] Moreover, learners' sensitivity to these nouns' stereotypical gender in English had been confirmed independently in a gender rating questionnaire. This makes it unlikely that the absence of early effects of the accessible antecedent's gender in the German reading time data reflected a lack of sensitivity to the stereotypical gender manipulation.

The present results are indicative of L1/L2 differences in the timing of syntactic coreference constraints during online reference resolution, with the application of binding condition A being delayed in L2 processing. What remains unclear, however, is precisely what factor or factors were responsible for the learners' initial focusing on the inaccessible antecedent. Possible candidates include the inaccessible antecedent's prominence in the preceding extrasentential discourse, or a preference for linking the reflexive to the current sentence's matrix subject. From a linguistic perspective, linking the reflexive to an extrasentential (or any other non c-commanding) antecedent involves discourse-based coreference assignment without binding, whereas linking it to the matrix subject might be argued to involve syntactic LD binding, which is contingent on c-command. To help us tease apart these two possibilities, we designed a second experiment using materials in which the inaccessible antecedent was structurally embedded and did not c-command the reflexive.

EXPERIMENT 2

Our follow-up experiment used materials similar to those used in Experiment 1 but with the inaccessible antecedent not c-commanding the reflexive. Felser et al.'s (2009) results suggest that L2 learners and native speakers are alike in that they do not take a non-c-commanding competitor antecedent into consideration during processing. Experiment 2 should thus help us determine whether, during their initial processing of the reflexive, learners attempt coreference assignment without binding, or whether they have an initial preference for LD binding.

Method

Participants. Participants included 26 German-speaking learners of L2 English (7 males, mean age = 24.8) and 28 native English speakers (16 males, mean age = 22.1), again recruited from the University of Essex community. None of the participants had any uncorrected visual impairments, and all were offered a modest fee to compensate them for the time and effort. Eighteen of the German speakers who took part in Experiment 1 also participated in Experiment 2. However, as there was a gap of between 2 and 5 months in between testing sessions, and because different filler items were used in each experiment, it is unlikely that prior exposure to similar experimental items had any effect on the L2 learners' performance in the second experiment. The learners had first been exposed to English regularly between ages 7 and 13 at school, and had been staying in the United Kingdom for 39.91 months on average ($SD = 60.25$ months) at the time of the experiment. They scored an average of 84.65% ($SD = 11.77\%$) correct in the Quick Placement

Table 6. *Native speakers (NSs) and nonnative speakers' (NNSs) mean (standard deviation) percentages of correct choices of the accessible antecedent per condition in the offline antecedent identification task in Experiment 2*

	Reflexive	Pronoun
NSs ($n = 28$)		
Single match	100 (0)	98 (7)
Double match	99 (5)	88 (2)
NNSs ($n = 26$)		
Single match	99 (3)	99 (3)
Double match	96 (9)	94 (14)

Test, which identified 18 of them as "upper advanced," five as "lower advanced," and three as "upper intermediate" learners.

As in Experiment 1, we carried out an offline antecedent identification task to confirm whether our participants were aware of the binding requirements of English reflexives, which was administered as a written questionnaire. The materials were parallel to those used in Experiment 1 and contained both reflexive and pronoun conditions. The reflexive conditions were modified so that the inaccessible antecedent was now contained within a subordinate (relative) clause and the accessible one in matrix subject position, as shown in (11a,b), whereas the pronoun conditions were the same as in (9c,d) above.

(11) a. REFLEXIVE, SINGLE MATCH
The grandfather that Emma was talking to explained himself carefully.
b. REFLEXIVE, DOUBLE MATCH
The grandfather that Adam was talking to explained himself carefully.

Again, there were very few "either" responses (native speakers = 4.23%, German learners = 0.96%). Table 6 shows the native and nonnative participants' mean percentages of correct antecedent choices. The learners performed about as well as the native speakers in this task, confirming that they knew the binding properties of English reflexives and pronouns.

Materials. For the eye-movement monitoring experiment, we created 24 sets of short texts parallel to those used in Experiment 1 but with the linear ordering of the accessible and inaccessible antecedents reversed, as shown in (12a–d) below.

(12) a. ACCESSIBLE MATCH, INACCESSIBLE MATCH
James has worked at the army hospital for years. The soldier that he treated on the ward wounded himself while on duty in the Far East. Life must be difficult when you are in the army.

b. ACCESSIBLE MATCH, INACCESSIBLE MISMATCH
Helen has worked at the army hospital for years. The soldier that she treated on the ward wounded himself while on duty in the Far East. Life must be difficult when you are in the army.
c. ACCESSIBLE MISMATCH, INACCESSIBLE MATCH
Helen has worked at the army hospital for years. The soldier that she treated on the ward wounded herself while on duty in the Far East. Life must be difficult when you are in the army.
d. ACCESSIBLE MISMATCH, INACCESSIBLE MISMATCH
James has worked at the army hospital for years. The soldier that he treated on the ward wounded herself while on duty in the Far East. Life must be difficult when you are in the army.

In the critical second sentences in (12a–d), the accessible antecedent *the soldier* is the matrix subject, whereas the structurally inaccessible one (*he/she*) is the subject of a relative clause modifying *the soldier*, so does not c-command the reflexive. This modification to the materials from Experiment 1 should allow us to empirically dissociate effects of the two antecedents' extrasentential discourse prominence and their intrasentential structural salience (subjecthood/c-command). Based on the results from Sturt (2003) and Xiang et al. (2009), we expect the native control group to show effects of the accessible antecedent's gender upon their first reading of the reflexive, in the absence of any reliable effects of the inaccessible one. If our nonnative participants initially try to link the reflexive to the current sentence's matrix subject (which happens to be the correct antecedent here), then their reading times should not be affected by the gender manipulation of the inaccessible antecedent, either, at least not in early eye-movement measures. If, in contrast, the learners initially attempt to interpret the reflexive via discourse-based coreference assignment, we should see early main effects of the inaccessible antecedent's gender also in this experiment, with effects of the accessible antecedent being again delayed.

Note that the relative surface ordering of the accessible and inaccessible antecedent has also changed, with the inaccessible one now linearly closer to the reflexive. However, on the assumption that the object relative clause contains a postverbal gap associated with the accessible antecedent (*The soldier that he treated __ on the ward* . . .), exactly how "distance" is to be defined here is not as obvious at it might seem.[6] We will return to this issue in our general discussion below.

The gender rating questionnaire from Experiment 1 was also administered to participants in Experiment 2. Again, participants from both groups rated the female biased nouns lower on the stereotype scale (2.54 and 2.69 for the native and L2 learner groups, respectively) than the male biased nouns (5.78 for both groups), and this difference was reliable: for the natives, $t_1\ (27) = 16.01, p < .001$; $t_2\ (23) = 20.92, p < .001$; for the L2 learners, $t_1\ (25) = 11.00, p < .001$; $t_2\ (23) = 22.83, p < .001$.

The experimental materials were distributed across four presentation lists, mixed with 50 new filler items and pseudorandomized. Two-thirds of all trials were again followed by a yes/no comprehension question that did not directly probe participants' interpretation of the reflexive.

Table 7. *Native speakers (NSs) and nonnative speakers' (NNSs) mean (standard deviation) reading times in the reflexive region (ms) in Experiment 2*

	1st Fixation Durations	1st Pass Reading Time	Regression Path Duration	Rereading Time
Accessible match, inaccessible match (=12a)				
NSs	210 (32)	226 (53)	266 (75)	135 (115)
NNSs	236 (32)	275 (63)	310 (100)	134 (121)
Accessible match, inaccessible mismatch (=12b)				
NSs	223 (46)	257 (67)	311 (125)	117 (93)
NNSs	248 (37)	298 (69)	371 (136)	143 (114)
Accessible mismatch, inaccessible match (=12c)				
NSs	220 (36)	252 (77)	324 (125)	173 (107)
NNSs	232 (44)	266 (65)	325 (104)	210 (167)
Accessible mismatch, inaccessible mismatch (=12d)				
NSs	219 (40)	237 (57)	298 (99)	163 (111)
NNSs	234 (47)	286 (80)	347 (99)	186 (161)

Procedure. The experimental, data trimming, and analysis procedures were the same as in Experiment 1.

Results

Participants' overall response accuracy to the end of trial comprehension questions was 90.21% for the native and 90.17% for the nonnative group, indicating that both paid attention to the task and read the stimulus materials for meaning.

Track loss accounted for 0.30% of the native speakers' data, and none of the L2 group's data. Vocabulary screening resulted in the removal of 0.96% of the German data. The critical reflexive region was skipped 14.3% of the time by the native speakers and 3.7% by the learners, and the postcritical region was skipped 6.7% and 0.7% of the time, respectively. Outlier removal of reading times 2.5 standard deviations beyond each participant's means per region and measure accounted for the removal of no more than 5.17% of the data. Table 7 and Table 8 provide summaries of the two participant groups' mean reading times per region of interest after data cleaning, for four eye-movement measures.

Preliminary mixed ANOVAs showed main effects of group in all measures at the reflexive region: first fixation durations, $F_1\ (1, 53) = 5.54, p < .05$; $F_2\ (1, 23) = 16.37, p < .01$; first-pass reading times, $F_1\ (1, 53) = 6.90, p < .05$; $F_2\ (1, 23) = 30.10, p < .001$; regression path durations, $F_1\ (1, 53) = 3.31, p = .075$; $F_2\ (1, 23) = 16.14, p < .001$; rereading times, $F_1\ (1, 53) = 0.57, p = .452$; $F_2\ (1, 23) = 4.67, p < .05$, reflecting slightly longer reading times in the German compared to the native group. There was also a marginally significant three-way interaction in participants' first-pass reading times, $F_1\ (1, 53) = 3.44, p = .069$; $F_2\ (1, 23) = 4.28, p = .050$. At the postcritical region, main effects of group were again seen in all measures: first fixation durations, $F_1\ (1, 53) = 6.80, p < .05$; $F_2\ (1, 23) = 19.19, p < .001$; first-pass reading times, $F_1\ (1, 53) = 20.20, p < .001$; F_2 (1,

Table 8. *Native speakers (NSs) and nonnative speakers' (NNSs) mean (standard deviation) reading times in the postcritical region (ms) in Experiment 2*

	1st Fixation Durations	1st Pass Reading Time	Regression Path Duration	Rereading Time
Accessible match, inaccessible match (=12a)				
NSs	222 (37)	302 (84)	374 (102)	124 (130)
NNSs	247 (39)	407 (99)	450 (120)	195 (207)
Accessible match, inaccessible mismatch (=12b)				
NSs	217 (39)	298 (73)	371 (121)	145 (102)
NNSs	243 (51)	413 (120)	510 (193)	174 (176)
Accessible mismatch, inaccessible match (=12c)				
NSs	231 (41)	326 (105)	481 (154)	175 (133)
NNSs	264 (59)	430 (120)	578 (163)	229 (243)
Accessible mismatch, inaccessible mismatch (=12d)				
NSs	224 (39)	316 (78)	485 (137)	206 (139)
NNSs	248 (51)	396 (106)	553 (144)	234 (280)

23) = 79.36, $p < .001$; regression path durations, F_1 (1, 53) = 10.24, $p < .01$; F_2 (1, 23) = 45.82, $p < .001$; rereading times, F_1 (1, 53) = 1.10, $p = .300$; F_2 (1, 23) = 28.29, $p < .001$, along with a marginal three-way interaction in regression path durations in the analysis by participants, F_1 (1, 53) = 2.88, $p = .096$; F_2 (1, 23) = 1.06, $p = .315$. Given this, and in order to keep the current analysis parallel to that of Experiment 1, we again proceeded to examine the two participant groups' reading time patterns separately.

Native speakers. Table 9 shows a summary of the ANOVA results for the two interest regions. At the reflexive region, we found reliable interactions between the factors accessible and inaccessible antecedent in first-pass reading times and regression path durations. Numerically, reading times were longer when one of the two potential antecedents matched and the other mismatched the reflexive's gender, compared to the double match (=12a) or double mismatch (=12d) conditions. For first-pass times, *t* tests showed differences between the two accessible match conditions (12a) and (12b): 226 versus 257 ms; t_1 (27) = 3.18, $p < .01$; t_2 (23) = 2.64, $p < .05$, between the two "inaccessible match" conditions (12a) and (12c): 226 versus 252 ms; t_1 (27) = 1.93, $p = .065$; t_2 (23) = 3.01, $p < .05$, and between the two inaccessible mismatch conditions (12b) and (12d) in the analysis by items: 257 versus 237 ms; t_1 (27) = 1.59, $p = .123$; t_2 (23) = 2.19, $p < .05$. For regression path durations, we found differences between conditions (12a) and (12b), marginal by items: 266 versus 311 ms; t_1 (27) = 2.55, $p < .05$; t_2 (23) = 1.73, $p = .096$, and conditions (12a) and (12c): 266 versus 324 ms; t_1 (27) = 2.95, $p < .01$; t_2 (23) = 2.69, $p < .05$. The expected main effect of the accessible antecedent, in the absence of any effects of the inaccessible one, was seen only during the native group's rereading of the reflexive region.

At the postcritical region, we found reliable main effects of the accessible antecedent for regression path durations and rereading times, with the accessible

Table 9. *Summary of analysis of variance results for the reflexive and postcritical regions in Experiment 2 for native speakers*

		Reflexive Region				Postcritical Region			
		1st Fixations	1st Pass Times	Regression Path	Rereading Times	1st Fixations	1st Pass Times	Regression Path	Rereading Times
Accessible	$F1$ (1, 27)	<1	<1	1.69	8.47**	1.57	3.33†	18.94**	9.09**
Antecedent	$F2$ (1, 23)	<1	<1	1.45	5.79*	2.07	2.48	16.33**	11.55**
Inaccessible	$F1$ (1, 27)	1.60	1.91	<1	1.69	1.81	<1	<1	2.03
Antecedent	$F2$ (1, 23)	1.24	<1	<1	<1	1.14	<1	<1	1.79
Interaction	$F1$ (1, 27)	1.84	8.45**	6.68*	<1	<1	<1	<1	<1
	$F2$ (1, 23)	1.73	11.92**	6.55*	<1	<1	<1	<1	<1

†$p < .1$. *$p < .05$. **$p < .01$.

mismatch conditions eliciting longer reading times than the accessible match conditions. There were no other significant main effects or interactions. Taken together, the above results suggest that the native speakers might have briefly experienced some competition between *both* potential antecedents when first encountering the reflexive, reflected in elevated reading times when one of them mismatched the reflexive in gender.

L2 learners. The nonnative group's ANOVA results are summarized in Table 10. At the critical reflexive region, the German group again patterned differently from the natives, and similar to the L2 group in Experiment 1, in that they tended to show longer reading times in the inaccessible mismatch compared to the inaccessible match conditions in early eye-movement measures. This was reflected in main effects of the inaccessible antecedent in first-pass reading times, marginal by participants, and in their regression path durations (shown in Figure 2), which suggest that the nonnative speakers initially tried to link the reflexive to the inaccessible antecedent. A main effect of the accessible antecedent was seen only in the L2 group's rereading times. There were no other significant main effects or interactions.

A significant main effect of the accessible antecedent was also seen in the regression path durations at the postcritical region, modulated by an interaction with the factor inaccessible antecedent in the analysis by participants. In this measure, the double match condition (12a) had numerically shorter reading times than the other conditions. Statistically, *t* tests revealed that the inaccessible mismatch conditions had longer reading times than the double match condition: (12a) versus (12c), 450 versus 578 ms; t_1 (25) = 5.68, $p < .001$; t_2 (23) = 2.94, $p < .01$; (12a) versus (12d): 450 versus 553 ms; t_1 (25) = 4.54, $p < .001$; t_2 (23) = 2.35, $p < .05$. Condition (12b) (accessible match, inaccessible mismatch) also had shorter reading times than Condition (12c) (accessible mismatch, inaccessible match): 510 versus 578 ms; t_1 (25) = 1.86, $p = .075$; t_2 (23) = 2.37, $p < .05$, but no other comparisons were reliable. No other significant effects or interactions were observed in the other reading time measures for this region.

Additional analyses of only the "upper advanced" learners" ($n = 18$) reading times on the reflexive revealed the same pattern as in our main analysis, with a main effect of the accessible antecedent restricted to rereading times. There is thus no evidence in our L2 data to suggest that the more advanced learners patterned with the native speakers here.

Discussion

The results from our follow-up experiment are similar to those from Experiment 1 in that the learners, unlike the native speaker controls, initially focused on the inaccessible antecedent only. As this was the case despite the accessible antecedent now being in matrix subject position, the results from the learners suggest that the crucial factor involved during early processing stages was the inaccessible antecedent's prominence in the (extrasentential) discourse, rather than its intrasentential structural salience. In other words, the L2 results suggest that the learners initially attempted discourse-mediated coreference assignment

Table 10. *Summary of analysis of variance results for the reflexive and postcritical regions in Experiment 2 for second language learners*

		Reflexive Region				Postcritical Region			
		1st Fixations	1st Pass Times	Regression Path	Rereading Times	1st Fixations	1st Pass Times	Regression Path	Rereading Times
Accessible	$F1$ (1, 25)	1.74	1.43	<1	12.24**	2.94†	<1	20.87**	3.07†
Antecedent	$F2$ (1, 23)	1.01	<1	<1	5.85*	2.40	<1	9.94**	1.94
Inaccessible	$F1$ (1, 25)	1.42	3.65†	4.38*	<1	3.04†	<1	<1	<1
Antecedent	$F2$ (1, 23)	1.19	5.41*	5.25*	<1	1.40	<1	<1	<1
Interaction	$F1$ (1, 25)	<1	<1	<1	<1	1.15	2.67	4.46*	<1
	$F2$ (1, 23)	<1	<1	<1	<1	<1	1.01	1.81	<1

†$p < .1$. *$p < .05$. **$p < .01$.

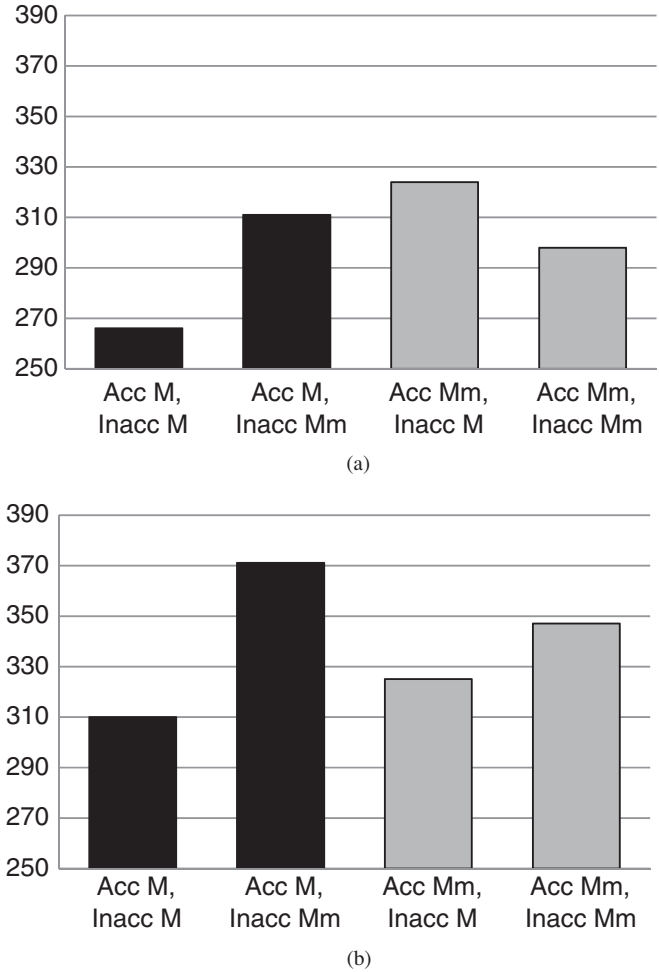

Figure 2. (a) Native and (b) nonnative speakers' mean regression path times (ms) on the reflexive in Experiment 2; Acc M, accessible match; Acc Mm, accessible mismatch; Inacc M, inaccessible match; Inacc Mm, inaccessible mismatch.

without binding. As in Experiment 1, reliable effects of the accessible antecedent's gender (our experimental diagnostic for the application of condition A) were not visible until the learners rereading of the reflexive region.

One difference to the results from Experiment 1 is that our native group also seemed to experience some degree of temporary interference by the inaccessible antecedent, as witnessed by the interaction seen in early eye movement measures. That is, the native group showed longer first-pass and regression path times in (12b) (*The soldier that she treated . . . injured himself*) than in (12a) (*The soldier*

that he treated . . . injured himself). This effect was fairly short-lived, however, and was not observed at a point in time before we also observed effects of the accessible antecedent, and was not seen in the native speakers' rereading times or their reading times at the postcritical region.

Note that in his second experiment, which used materials similar to ours, Sturt (2003) actually found a marginal trend toward an effect of the inaccessible antecedent's gender during participants' initial reading of the reflexive region, a trend consistent with the results from our L1 group. Similarly, Xiang et al. (2009) observed a nonsignificant trend toward a late positivity elicited by a mismatching inaccessible antecedent (*The well-known surgeon . . . Jennifer . . . himself*) compared to a matching one, suggesting that the inaccessible antecedent may not have been completely ignored here, either. Together with our findings, these observations indicate that even though condition A applies immediately in native sentence processing, the presence of inaccessible competitor antecedents may sometimes lead to greater processing cost.

The results from this experiment will be discussed in more detail below, together with the results from Experiment 1.

GENERAL DISCUSSION

Taken together, the results from the current study add to the growing body of research that has revealed L1/L2 differences in the processing of discontinuous grammatical dependencies (Clahsen & Felser, 2006; Dallas & Kaan, 2008). In both eye-movement experiments, our learners behaved differently from the native controls in that they initially tried to link the reflexive to a structurally inaccessible competitor antecedent, with their consideration of the accessible antecedent consistently being delayed. The L1/L2 performance differences we observed cannot obviously be accounted for by factors such as L1 influence, slower processing speed, or insufficient grammatical knowledge.

First, recall that argument reflexives in our learners' native language, German, are like English reflexives in that they require syntactic binding by a local antecedent. The learners' initial preference for a nonlocal (Experiment 1) or non-c-commanding (Experiment 2) antecedent can thus hardly be explained by negative L1 transfer. Second, note that the L1/L2 differences in general reading speed in our study were very small, with the learners not slowed down by having to read in a foreign script, as was arguably the case for the Japanese participants in Felser et al.'s (2009) study. There is thus no evidence in our results to suggest that the observed L1/L2 differences merely reflect slower processing speed. Rather than showing nativelike reading-time patterns that were temporally delayed, our native and nonnative participant groups showed qualitatively different patterns across the experimental conditions in both experiments. Third, recall that our nonnative participants were mostly advanced learners who had been immersed in English for an average period of about 3 years at the time of testing. Their performance in our offline questionnaire tasks was nativelike, so that there is no reason to suspect that they were unaware of the binding properties of English reflexives, or that they should have mistaken reflexives for nonreflexive pronouns.

As the principal aim of our study was to examine the role of structural versus discourse-level coreference constraints on anaphor resolution, let us now take a closer look at how these interact in real time during L1 and L2 processing.

The application of binding condition A in nonnative sentence processing

Our study is the first to examine the timing of condition A in nonnative sentence processing. Unlike what has been reported for native speakers in previous L1 processing studies (Nicol & Swinney, 1989; Sturt, 2003; Xiang et al., 2009) and corroborated by the results from our native speaker controls, our L2 learners clearly violated condition A during early processing stages. Their offline interpretation of English reflexives, in contrast, was nativelike. That is, rather than being insensitive to binding condition A, it appears that our nonnative participants merely took longer than the native controls to apply phrase structure-based coreference constraints during processing, after temporarily considering a structurally inaccessible competitor antecedent.

The results from our German-speaking participants are consistent with those reported by Felser et al. (2009) for Japanese-speaking learners of English. Given their finding that their learners initially considered a nonlocal matrix subject as a possible antecedent for a reflexive but not a non-c-commanding inaccessible antecedent, Felser et al.'s (2009) results could be taken to suggest that learners may violate the locality but not the c-command requirement of English reflexive binding. Our results indicate that this is not the case, however. The German-speaking participants in the current study were not influenced by c-command, at least not to any measurable extent. The combined results of Experiments 1 and 2 show that they initially favored the inaccessible antecedent regardless of whether or not it c-commanded the reflexive. Thus, if we reconsider Felser et al.'s findings in the light of our current results, an explanation in terms of the inaccessible antecedent's relative discourse salience seems more likely. This issue will be addressed in the next section.

Structural versus discourse-level constraints

Recall that online anaphor resolution is influenced by a number of potentially interacting constraints, including morphological, syntactic, semantic, and discourse-level constraints. The present study has focused on only a subset of these, with the aim of determining how structural coreference constraints (notably, binding condition A) interact with discourse-level constraints on anaphor resolution. Discourse-level constraints on anaphor resolution include factors such as distance, first mention, subject- or topic-hood, parallelism, or pragmatic factors such as implicit causality, all of which may affect a potential antecedent's relative "accessibility" (compare Koornneef, 2008, pp. 47–51).

In Experiment 1, the accessible antecedent was favored both by binding condition A and by virtue of being the linearly closest potential antecedent. The inaccessible antecedent, in contrast, was favored by several discourse-level constraints including first mention, being mentioned twice rather than only once, and by being a matrix subject (a factor that coincides with c-command, of course). In

Experiment 2, that the accessible antecedent now served as the matrix subject did not prevent the nonnative speakers from initially considering only the inaccessible antecedent. This suggests that factors such as first mention and/or frequency of mention weighed more strongly than a potential antecedent's sentence-internal structural salience here.

If we were to consider the results from Experiment 2 on their own, it might seem conceivable that the learners tried to link the reflexive to the linearly closest potential antecedent (the pronoun *he* or *she*, referring back to the inaccessible antecedent). Regarding the native speakers, the interactions seen in some of the early reading time measures might have been due to binding condition A interacting with linear distance during early processing stages. However, observe that distance could potentially be computed in two different ways here, due to the relative clause containing an object gap associated with the accessible antecedent. If the relativized noun phrase *the soldier* is mentally reactivated when participants encounter the embedded transitive verb *treat*, as illustrated in (13) below, then this would effectively render a representation of the accessible antecedent linearly closer to the reflexive, relative to the inaccessible antecedent.

(13) [*The soldier*] that she treated [*the soldier*] on the ward wounded himself . . .

The reactivation of fronted or relativized constituents at gap sites has frequently been demonstrated in L1 processing research (e.g., Love & Swinney, 1996). That is, only if we define distance in terms of the surface word order does this factor favor the inaccessible antecedent in Experiment 2. If we take into account the likely mental reactivation of the accessible antecedent at the embedded object gap, in contrast, then distance actually favors the accessible antecedent here. Irrespective of this rather interesting ambivalence, recall that the results from Experiment 1 demonstrate quite clearly that linear proximity did not measurably affect the L2 learners' initial attempt at reference resolution.

The overall picture that emerges, then, shows that our learner groups' initial processing of reflexives was affected primarily by discourse-level factors, whereas the native speaker controls showed sensitivity to structural coreference constraints from the earliest stages in processing onward. The results from our nonnative participants indicate that the timing of discourse-level constraints precedes that of phrase structure-based coreference constraints in L2 processing. Our L2 data are thus not consistent with serial "syntax first" models of sentence processing. In particular, our finding that L2 learners seem initially to attempt interpreting reflexives via coreference assignment rather than binding does not support the hypothesis that in anaphor resolution, the discourse-based coreference route will only be considered where syntactic binding is unavailable (Koornneef, 2008; Reinhart, 1983).

Our results are compatible, however, with processing models that assume the existence of multiple processing pathways potentially operating in parallel (e.g., Ferreira, Bailey, & Ferraro, 2002; Townsend & Bever, 2001), and the hypothesis that structural processing is compromised in nonnative comprehension (Clahsen & Felser, 2006). From the point of view of Clahsen and Felser's shallow structure hypothesis for L2 processing, phrase structure-based principles such as binding

condition A cannot be applied immediately, even by speakers whose native language is similar to the L2 in relevant respects, because mental representations encoding the syntactic configurations over which these are defined cannot be assembled quickly enough, or lack sufficient detail. Instead, the nonstructural (i.e., semantics and discourse-based) processing route may be the faster one in nonnative anaphor resolution, with learners initially influenced mainly by discourse-level factors and the application of structural coreference constraints delayed.

CONCLUSION

We set out to further explore the information sources that guide real-time anaphor resolution in nonnative language processing. Using eye-movement recording during reading allowed us to tap into online comprehension processes that are not otherwise open to direct observation. Our results revealed clear L1/L2 differences in the relative timing of structural versus discourse-level constraints, with effects of a discourse-prominent but structurally inaccessible antecedent temporally preceding those of the structurally accessible one in L2 processing. Our finding that discourse-level factors may lead nonnative speakers even from non-LD binding backgrounds to temporarily violate binding condition A suggests not only that L2 processing is guided more strongly by nonstructural cues to interpretation than native language processing but also that the nonstructural processing route may be faster in L2 processing. Further examination of the role and relative timing of discourse-level information in nonnative compared to native language processing might provide a fruitful topic for future research.

ACKNOWLEDGMENTS

This research was supported by Economic and Social Research Council Grant RES-000-22-2508 (to C.F.). We thank Harald Clahsen, the audiences at the 19th Annual Conference of the European Second Language Association, Cork, and the Generative Approaches to Language Acquisition 2009 Conference, Lisbon, and three anonymous *Applied Psycholinguistics* reviewers for stimulating discussion and constructive comments.

NOTES

1. We will use the term *L2 learner* here to refer to people who started learning another language after early childhood, and after acquiring the core properties of their native language(s).
2. The situation is less clear for reflexives in so-called "picture" noun phrases containing possessors such as *Bill's picture of himself*. The results from a visual-world eye-movement study reported by Runner, Sussman, and Tanenhaus (2003) showed that this kind of reflexive also allowed nonlocal binding, which according to the authors, can be accounted for by assuming that they are in fact logophors rather than argument reflexives.
3. Evidence that native speakers may consider an inaccessible antecedent during processing has also been reported by Badecker and Straub (2002), who found that a structurally

and discourse-prominent but inaccessible antecedent such as *John* in sentences like *John thought that Bill owed himself another opportunity to solve the problem* affected participants' reading times two words down from the reflexive.

4. Complete sets of the experimental items used in Experiments 1 and 2 can be made available upon request by the first author.
5. Exceptions include the foreign words *Babysitter* and *Cheerleader*, whose citation forms carry masculine gender in German (determined by the *-er* nominalization suffix), and *Model*, which carries neuter gender. German speakers nevertheless perceive these as stereotypically female occupations.
6. The same reasoning applies to the materials used in Xiang et al.'s (2009) study, so that the absence of reliable effects of the inaccessible antecedent in Xiang et al.'s ERP experiment cannot unambiguously be attributed to the workings of condition A.

REFERENCES

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.

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.

Dallas, A., & Kaan, E. (2008). Second language processing of filler-gap dependencies by late learners. *Language and Linguistics Compass, 2*, 372–388.

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

Demirci, M. (2001). Acquisition of binding of English reflexives by Turkish L2 learners: A neo-gricean pragmatic account. *Journal of Pragmatics, 33*, 753–775.

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*, 485–502.

Ferreira, F., Bailey, K., & Ferraro, V. (2002). Good enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.

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.

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.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research, 22*, 369–397.

Koornneef, A. (2008). *Eye-catching anaphora*. LOT Dissertation Series 90. Utrecht: Netherlands Graduate School of Linguistics (LOT).

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*. Somerville, MA: Cascadilla Proceedings Project.

Love, T., & Swinney, D. (1996). Coreference processing and levels of analysis in object-relative constructions: Demonstration of antecedent reactivation with the cross-modal priming paradigm. *Journal of Psycholinguistic Research, 25*, 5–24.

McDonald, J. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgement performance by late second language learners. *Journal of Memory and Language, 55*, 381–401.

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*. Oxford: Blackwell.

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

Reinhart, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.

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.

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.

Staub, A., & Rayner, K. (2007). Eye movements and online comprehension processes. In M. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*. 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.

Townsend, D., & Bever, T. (2001). *Sentence comprehension: The integration of habits and rules.* Cambridge, MA: MIT Press.

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

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