Published in*: Journal of Memory and Language* 65 (2011) 128–144.

Children's Processing of Reflexives and Pronouns in English: Evidence from Eye-movements During Listening

Kaili Clackson **Claudia Felser** **Harald Clahsen**

(University of Essex)

Running Head: *Children's Processing of Reflexives and Pronouns*

Corresponding Author:
Harald Clahsen
Department of Linguistics
University of Essex
Colchester, C04 3SQ, UK
harald@essex.ac.uk

Abstract

This study examined how six-to-nine-year-old English-speaking children and adults establish anaphoric dependencies during auditory sentence comprehension. Using eye-movement monitoring during listening and a corresponding sentence-picture judgment task, we investigated both the ultimate interpretation and the online processing of reflexives in comparison to non-reflexive pronouns, focusing on how binding constraints interact with a competitor antecedent's relative (discourse) prominence. Whilst our offline results show that the children's ultimate interpretation for reflexives was constrained by binding principles in the same way as adults', the eye-movement data revealed that during processing, children were temporarily more distracted than adults when multiple cues supported a prominent competitor antecedent. These results indicate that in addition to binding principles, children's online referential decisions are also affected by discourse-level information. We suggest that the observed child/adult differences stem from children's greater difficulty, compared to adults, in controlling multiple sources of information during sentence comprehension.

Introduction

A well-established finding from previous research on child language acquisition concerns an asymmetry in children's offline referential interpretations of reflexives and (non-reflexive) pronouns. Several studies on languages such as English, French, and Dutch using offline picture-matching, truth-value judgment, and act-out tasks found that from the age of around three years, children interpret reflexives in an adult-like manner, whereas their interpretation of pronouns remains non-adult-like until around six years of age (see Guasti, 2002, for a review). Thus, young children correctly interpret reflexives such as *herself* in (1a) as coreferential with the closest (= local) antecedent (i.e., with *Susan* in (1a)), in accordance with Principle A of the Binding Theory (Chomsky, 1981), but sometimes also incorrectly accept a similar reading for (1b), allowing *her* to be coreferential with *Susan*, thereby violating Principle B.

(1) a. Jane_1 says that [Susan_2 hurt $\text{herself}_{*1/2}$]

b. Jane_1 says that [Susan_2 hurt $\text{her}_{1/*2}$]

This phenomenon is known as the *delay of Principle B effect* in the acquisition literature (Guasti, 2002: 296). Several explanations have been proposed for this asymmetry (e.g., Thornton & Wexler, 1999; Reinhart, 2006; Matthews, Lieven, Theakston & Tomasello, 2009; van Rijn, van Rijn & Hendriks, 2010). What is common to most accounts is that the interpretation of reflexives is thought to be uniquely determined by a structurally defined constraint (Principle A) which forms part of a child's grammar early on (but see Matthews et al., 2009, for an alternative view). Binding Principle A requires argument reflexives to be bound by a local antecedent in English, while Principle B states that (non-reflexive) pronouns must

remain locally unbound (Chomsky, 1981). Consequently, Principle B does not determine a unique referent for a pronoun but merely rules out local binding. The interpretation of pronouns thus requires additional knowledge and/or computation, and recourse to other information sources, which may not yet be available or adult-like in young children. Adults also sometimes permit coreferential interpretations with local antecedents for pronouns, but only under special circumstances, e.g., when the pronoun is stressed (Thornton &Wexler, 1999) or when a coreferential interpretation is distinct in meaning from a bound variable interpretation (Reinhart, 2006). Young children, however, do not consider these special circumstances in the same way as adults and over-accept local coreference interpretations for pronouns.

The *time-course* of reflexive anaphor resolution during sentence processing has so far mainly been studied for adults. Studies using time-sensitive measures such as eye-movements or event-related potentials have shown that adult comprehenders adhere to Binding Principle A during processing, quickly linking reflexives to their corresponding antecedents (see, among others, Nicol & Swinney, 1989; Harris, Wexler & Holcomb, 2000; Xiang, Dillon & Phillips, 2009). Although early studies using the cross-modal priming technique (Nicol, 1988; Nicol & Swinney, 1989) suggested that during anaphor resolution, the adult parser only considers syntactically appropriate (henceforth, 'accessible') antecedents but not structurally inappropriate (henceforth, 'inaccessible') ones, the 'binding-as-initial-filter' hypothesis has more recently been qualified, by showing that syntactically and pragmatically salient but inaccessible antecedents can also affect online anaphor resolution in monolingual adults (Badecker & Straub, 2002; Cunnings & Felser, in press; Sturt, 2003). Sturt (2003), for example, found in a series of eye-movement-during-reading experiments that adult native speakers' initial fixations on a reflexive were not affected by the

gender of a structurally inaccessible antecedent. In later (second-pass) reading time measures, on the other hand, a mismatch in gender between the reflexive and a discourse-prominent but inaccessible antecedent led to elevated reading times compared to a gender-matching inaccessible antecedent. Taken together, these findings suggest that in addition to structurally-determined antecedents, the adult parser may also consider highly prominent competitor antecedents for argument reflexives, specifically during later stages of processing.

For children, several studies have used time-sensitive techniques to show that children's online comprehension of pronouns is affected by lexical and discourse-level information such as gender and discourse prominence (e.g., Song & Fisher, 2005, 2007; Arnold, Brown-Schmidt & Trueswell, 2007; Pyykkönen, Matthews & Järvikivi, 2010). Few previous studies have directly compared the processing of pronouns and reflexives in online experiments, however. Using a cross-modal (picture) priming task with four-to-six-year-old English-speaking children, McKee, Nicol, and McDaniel (1993) found priming effects of the local antecedent for reflexives in sentences such as *The alligator knows that the leopard with green eyes is patting himself/him on the head with a pillow,* that is, priming by the accessible referent as determined by Binding Principle A (i.e., the leopard). For pronouns, on the other hand, they did not find any priming for local antecedents, which are inaccessible according to Principle B. According to McKee et al., these results provide evidence for children employing structural information, i.e., Binding Principles A and B, during online processing. However, as McKee et al. only tested for priming of the local antecedent, we cannot be sure what, if any, priming effects might have been found for a potential non-local competitor antecedent.

Love, Walenski and Swinney (2009) later used the materials from McKee et al.'s (1993) study to test for priming effects by the non-local antecedent (i.e., the alligator). They found that five-to-thirteen-year-old children showed significant priming of the accessible antecedent for pronouns if the auditory stimuli were presented at a normal speech rate, but not if the speech input was slowed. In a complementary offline sentence/picture matching task, on the other hand, slowing the speech rate improved the children's performance on pronouns (see also van Rijn et al., 2010). For stimuli presented at a normal speech rate, children's offline accuracy scores for pronouns were found to be modulated by age, with children above the age of eight performing almost perfectly but younger ones performing at chance-level. According to Love et al., the observed offline/online differences suggest that automatic structural processing routines are adult-like by age four or five (but can be disrupted by slowing down the speech rate), whereas children's ability to use the kind of metalinguistic knowledge required for pronoun interpretation in offline tasks is developmentally delayed.

In another study, Sekerina, Stromswold and Hestvik (2004) employed the eye-monitoring-during-listening technique, which avoids the need for any meta-linguistic task and indicates changes in processing preferences over time; see Trueswell (2008) for a review. Sekerina et al.'s experimental stimuli included a lead-in sentence ('preamble') followed by a question, as shown in (1) below, and a two-picture visual display showing alternative interpretations of either a reflexive or a so-called 'short-distance' pronoun (which are known to be exempt from Principle B).

(1) PREAMBLE: In these pictures, you see a boy, a man, and a box. The boy has placed the box on the ground.

QUESTION: Which picture shows that the boy has placed the box behind *himself*/*him*?

Whilst four-to-seven-year-old children's ultimate picture choices revealed a strong preference for a sentence-internal antecedent for both reflexives and pronouns, Sekerina et al. found a contrast between the two types of anaphor in their eye-movement patterns, with reflexives eliciting a clear preference for the local referent (i.e, the boy) and pronouns showing relatively more looks to the picture in which the box was located behind the sentence-external referent (i.e., the man). This was the case for both children and adults, although the children only showed effects of the referential ambiguity of pronouns such as *him* in (1) around 1000ms later than adults. According to Sekerina et al., these results indicate that apart from requiring more time for accessing potential discourse referents and having more difficulty revising initial interpretation preferences, children process ambiguous pronouns in the same way as adults.

Note, however, that in Sekerina et al.'s materials, the critical sentences only contained a single potential antecedent for the reflexive or pronoun, whereas the competitor antecedent was mentioned only in the preamble. Thus, although the observed contrast in Sekerina et al.'s eye-movement data indicates that a sentence-external discourse referent is more likely to be considered for an ambiguous pronoun than for a reflexive, it is not clear from the results of this study whether children also process reflexives differently from pronouns when two competitor antecedents are available in the same sentence. Further studies on the time course of anaphor resolution in children are necessary to address this question. This is specifically the case for children's online processing of reflexives, for which there is much less

evidence from time-sensitive measures than for pronouns. The present study examines whether and when during processing children experience competition between structurally accessible and inaccessible antecedents for reflexives and pronouns, and how children's processing patterns differ from those of adults.

The Present Study

To investigate and compare the processes involved in children and adults' interpretation of reflexives and pronouns, we employed the eye-monitoring-during-listening technique, in which a participant's eye-movements are monitored while s/he is listening to spoken language. This technique provides detailed time-course sensitive measures on moment-to-moment language processing and is suitable for both children and adults (Trueswell, 2008; Fernald, Zangl, Portillo & Marchman, 2008). In addition, we determined referential decisions for reflexives and pronouns in children (and an adult control group) using an offline sentence-picture judgment task.

Given previous findings from offline studies, we expect six-to-nine-year-old children to be adult-like in their ultimate interpretations of reflexives but not necessarily of pronouns. There is, however, some evidence suggesting that the ability to dynamically control multiple sources of information during sentence processing is not yet fully developed in children at this age (e.g., Trueswell, 2008; Felser & Clahsen, 2009). If school-age children are indeed less efficient than adults at accessing and integrating multiple sources of information, then children's online processing may still not be fully adult-like even for reflexives. We may, for example, find that children require more time and/or find it more difficult than adults to deactivate potential (but structurally inaccessible) competitor antecedents for reflexives when making referential decisions during processing. Likewise, children's

processing of pronouns may also be different from adults', particularly for those children whose ultimate interpretations still (incorrectly) permit coreference between a local antecedent and a pronoun ('delay of Principle B').

Participants

Forty children were recruited from private and state primary schools in the Cambridge (UK) area who were either in year 2 (aged 6-7) or year 4 (aged 8-9) at the time of testing. All had acquired English as their native language from birth and were identified by their teachers as having no language or learning difficulties, and to be performing within normal parameters for their age. Parental consent was obtained prior to their testing. The children had a mean age of 8;0 (age range 6;3 to 9;9, 22 girls, 18 boys). They all took part in the eye-movement experiment and the offline judgment task.

The eye-movement experiment was also administered to a control group of 40 adult native speakers of English recruited from among the students and staff of the University of Essex, who were paid a small fee for their participation. Data from one adult participant could not be analysed, due to an error in sound recording, leaving 39 participants in the adult group (mean age: 21;3, age range: 18 to 39, 27 women, 12 men). The offline task was also performed with an adult native speaker control group (n=8, mean age: 28;6, age range: 19 to 34, five women, three men), all of whom were university educated and who did not take part in the main eye-movement experiment.

Experiment 1: Sentence-Picture Judgments

In order to assess how the children tested in the main (online) experiment ultimately interpret reflexives and pronouns, they took part in an offline sentence-

picture judgment task. Following the results from previous studies (e.g., Thornton & Wexler, 1999), we expect all children to perform adult-like on reflexives, in accordance with Binding Principle A, but that some children might over-accept local coreference interpretations for pronouns, in violation of Principle B.

Methods

The procedures and materials used for this task were adopted from van der Lely & Stollwerck (1997), based on Chien and Wexler's (Chien & Wexler, 1990) experiment 4. Participants saw a picture that either matched the contents of a question spoken by the experimenter (requiring a *yes* response) or did not match (requiring a *no* response). The task included two parallel sets of visual and auditory materials for reflexives and for pronouns. Each set consisted of five items; an example set for reflexives is shown in (2). For both sets of materials, the gender of the inaccessible antecedent was manipulated yielding two conditions, a '*double-match*' condition in which the pictures contained characters that were either all female or all male (2a), and a '*single-match'* condition (2b), with one female (e.g., *Kanga*) and one male character (*Christopher Robin*) of which only the correct antecedent matched in gender with the reflexive or pronoun.

(2) a. Double-match

This is *Christopher Robin*, this is *Pooh Bear*. Is *Pooh Bear* scratching *himself*?

Picture stimuli requiring: (i) *yes* response (ii) *no* response

b. Single-match

This is *Christopher Robin*, this is *Kanga*. Is *Christopher Robin* scratching *himself*?

Picture stimuli requiring: (i) *yes* response (ii) *no* response

The introductory sentence where the experimenter pointed to the two characters was included to make it more pragmatically appropriate for either character to then be referred to by a pronoun or reflexive. The experimenter spoke the introductory sentence, followed by a question to which the children had to reply *yes*/*no* respectively. Test questions used five action verbs (*tickle, scratch, pinch, point, wash*), with a different set of cartoon characters used to exemplify each verb in the picture materials, thus leading to five action-picture sets. Each sentence was presented twice in each of the conditions illustrated in (2), once with a picture that concurred with the contents of the sentence (requiring a *yes* response) and once with a picture that did not (requiring a *no* response). Eight filler items using similar pictures and including the quantifier *every* were included (e.g., *Is every monkey tickling himself?*), four with reflexives and four with pronouns. The resulting 48 items were presented in a pseudo-randomised order such that no consecutive items involved the same characters, or were from the same condition; and no more than three items requiring the same response (*yes* or *no*) were presented consecutively. Two further filler items were used as practice items.

Children were tested individually in a quiet room as part of the testing session which also included the eye-movement experiment, which was completed before the offline task was administered. The child sat at a table with the experimenter who explained that they were going to look at some cartoon pictures and that for each

picture the child would be asked a question and should answer *yes* or *no*. Children were given the option of taking a short break after the first half. Administering the judgment task took approximately ten minutes.

Results

The adults performed at 100% correct in all conditions, for both reflexives and pronouns. For the child data, mean percentages of correct (*yes* and *no*) responses for the two conditions are shown in Table 1, separately for reflexives and pronouns.

//INSERT TABLE 1 ABOUT HERE//

For reflexives, the children achieved high accuracy scores of 97% and above in both conditions and both for trials requiring a *yes* and for trials requiring a *no* response. For pronouns**,** the children again scored highly in the single-match condition, whereas their accuracy scores were lower in the double-match condition where both characters had the same gender. This was particularly the case when children were required to give a *no* response, in which case accuracy was significantly lower than in the corresponding single-match condition (86.5% vs. 98%, $t_1(39) = 4.16$, $p<.001$; $t_2(4) = 2.23$, $p = .090$). Further analyses of this condition showed that at an individual participant level, 20 children scored 100% correct for *no* responses in the double-match condition, while the remaining children occasionally responded incorrectly, thereby reducing the mean overall accuracy for pronouns, although the children's performance in this condition was not significantly correlated with age ($r = .19$, $p = .25$). The errors were distributed over all the verbs used.

These results confirm previous findings from offline tasks indicating a contrast between reflexives and pronouns in children's performance. Whilst all the children tested were able to reliably identify the correct referent for a reflexive, half of them occasionally accepted local coreference interpretations for pronouns. Thus children's ultimate referential interpretations for reflexives were adult-like and in accordance with Principle A, and not affected by a gender-matched but inaccessible competitor antecedent. Pronouns, however, were sometimes interpreted as being coreferential with the local antecedent (contra Principle B), specifically in cases without disambiguating gender cues.

Experiment 2: Eye-Movement Experiment

The purpose of this experiment was to uncover the time-course of listeners' referential decisions for reflexives, in comparison to pronouns. Participants listened to a series of short two-sentence paragraphs containing reflexives or pronouns whilst their eye-movements to visual displays containing pictures of potential referents were monitored. As the children's final interpretations for reflexives were adult-like, experiment 2 specifically examines whether this also holds for their ongoing referential decisions during sentence processing. If McKee et al.'s (1993) and Love et al.'s (2009) results from cross-modal priming generalize to other tasks, one would expect that the six-to-nine-year olds we tested should process both reflexives and pronouns in the same way as adults, despite their relatively poorer performance on pronouns in experiment 1. If, on the other hand, children's ability to handle multiple types of information during processing is not yet fully developed at this age (e.g., Trueswell, 2008), then an alternative outcome might be that children are temporarily more distracted by competitor antecedents for reflexives and/or pronouns than adults.

Materials

The materials consisted of spoken pairs of sentences and accompanying visual displays. Each display contained four pictures: two animate characters and an inanimate object mentioned in the sentences, and an inanimate distracter object not mentioned in the spoken sentences. Each experimental auditory stimulus contained two sentences involving two characters from the set of *Susan, Peter, Mr. Jones* and *Mrs. White.* The first sentence introduced the first character and established a felicitous context for the second sentence, which included the second character and the critical pronoun or reflexive. In each trial, the introduction of the second character was separated from the pronoun or reflexive by a phrase of 10-13 syllables which included the introduction of the inanimate object so as to direct participants' gaze away from either of the two characters to the picture of the inanimate object before the onset of the reflexive or pronoun. The stories were constructed specifically for use with children, with characters and objects familiar to school-age children. The auditory stimulus set comprised 24 experimental items each for reflexives and pronouns, each appearing in the same two conditions as in experiment 1, as illustrated in (3a,b). A full list of experimental auditory materials is provided in Appendix A.

(3) a. Double-match

Peter was waiting outside the corner shop. He watched as *Mr. Jones* bought a huge box of popcorn for *himself/him* over the counter.

b. Single-match

Susan was waiting outside the corner shop. She watched as *Mr. Jones* bought a huge box of popcorn for *himself/her* over the counter.

In addition to the experimental trials, the stimulus presentation lists for the adults included 48 auditory filler trials comprising a range of different grammatical constructions. Half of the filler trials involved the same characters as the experimental trials and the other half additional characters (queen, king, doctor, nurse). The children's presentation lists included a subset of 28 filler items, 12 with the same characters as in the experimental trials, and 16 with additional characters.

Finally, all experimental and filler trials were followed by a simple spoken *yes/no* comprehension question such as *Did Mr. Jones buy some popcorn?* to ensure that participants attended to the task and that any participants who did not could later be identified. The comprehension questions did not probe the interpretation of the critical reflexive or pronoun, so as to help ensure that participants' eye-movements reflected natural processing patterns and no specific attention was drawn to the focus of the experiment. Half the comprehension questions required a *yes* answer, and half a *no* answer, and these were equally split between the experimental and filler items. Sound files of all trials were recorded by two female native English speakers, one speaking the sentence pairs and one speaking the comprehension questions.

For each auditory trial, two visual displays were constructed, corresponding to the double-match and single-match conditions. Each visual display contained four pictures, as illustrated in Appendix B for the example trials shown in (3). The four pictures were positioned in the corners of the screen, with a small cross in the centre. The positioning of the pictures of the characters and the inanimate objects in the visual displays was counterbalanced across items. All pictures were black-and-white line drawings, of approximately the same size, and were not noticeably different in terms of visual saliency. All pictures were selected from a set of 520 pictures for which different norms are available from the International Picture Naming Project

(http://crl.ucsd.edu/~aszekely/ipnp/) with respect to their 'visual recognisability', which is expressed in the IPNP norms as the percentage of people who quickly identified a given picture. For the selected picture stimuli, the mean percentage from the IPNP norms was 97% (SD: 6%, range: 80-100%), indicating that the stimuli were easily recognizable.

Experimental trials with reflexives were presented in the same testing session as those with pronouns. Experimental trials were arranged in four lists according to a Latin Square design such that each participant saw each trial in only one condition (double-match or single-match), and each participant saw twelve trials with reflexives (six per condition) and twelve with pronouns (six per condition). The same set of filler trials was used with each list. All trials were presented in a pseudo-randomised order such that no more than two experimental trials occurred consecutively, and no more than three consecutive trials required the same (*yes* or *no*) response. The four lists were then reversed to create eight lists altogether. This was done to counteract any effects of fatigue so that items which were seen early on in the experiment by one participant were seen late in the experiment by another participant.

Procedure

Children were tested in a dedicated room at their school; adults in the visual world laboratory at the University of Essex. A consent form was signed and a short personal information questionnaire filled in by the adult participants or the children's guardians. Participants sat in a chair facing a projection screen on which the visual displays were projected, listened to the auditory stimuli through headphones, and responded to comprehension questions by pressing buttons on a gamepad. During the experiment participants' eye-movements were recorded by a SONY DSR-PD170P

digital camcorder recording 25 frames per second (i.e., one frame every 40ms) which was set up below the projection screen and trained on the participant's face. Children sat 1.5m from the screen, on which the display measured approximately 120x90cm, while adults sat 2m away from a projected visual display of 170x120cm. This ensured that when the video was played back, participants' eye-movements between pictures were distinct enough to be clearly interpreted. The presentation of visual and auditory stimuli was programmed using DMDX (Forster & Forster, 2003), and the sound output from the computer was split, with one line going to the headphones worn by the participant, and another going directly to the video camera, to ensure that the sound recorded by the video camera was exactly synchronized with what the participant heard.

At the start of each trial a small cross appeared in the centre of the screen for 2000ms followed by the visual display. After the visual display had been on the screen for 1000ms, the auditory stimulus began. The visual display remained on the screen during the auditory stimulus and disappeared when the participant responded to the comprehension question by pressing a button on a gamepad. The participant's response served to initiate the next trial.

Participants were first familiarized with the experimental equipment and their tasks, including the introduction of each of the animate characters used in the auditory materials along with his/her picture. This was followed by five practice trials, which served to further familiarize participants with the animate characters. Participants were asked to focus on the cross at the start of each trial, but to look where they liked on the screen once the visual display was shown. After the practice session participants were given the opportunity to ask questions before commencing the eye-movement experiment. During the experiment the researcher monitored the

participants' attention, gently reminding them to keep looking at the screen if their attention wandered. Participants were given a break after every 13 to 18 trials.

Data Coding and Analysis

Footage from the video camera was analysed using ELAN annotation software (Brugman & Russell, 2004). From the onset of the critical reflexive or pronoun, participant's gaze direction was recorded every frame for 2000ms (50 frames in total). Additional coding of the first 3000ms after the critical words for a subset of participants confirmed that the main differences between conditions appeared with the first two seconds of each trial. For each frame (every 40ms), the still image was inspected to determine the direction of gaze. A target was counted as 'fixated' for every frame where eyes were directed towards that picture. To ensure that coding was not influenced by the coders' expectations, gaze direction was initially coded as being towards the top left, top right, bottom left, bottom right, centre, or other (i.e., off-screen, or blinking), without the coder knowing to which particular pictures these directions related. Recorded gaze directions were then re-coded with reference to the visual display to show whether the participant was looking at the accessible antecedent, the inaccessible antecedent, the object, the distracter, the centre, or off-screen. Off-screen looks (which accounted for 6.6% of the total dataset) were treated as missing data. In order to assess the accuracy of the coding, data from ten adult and ten child participants selected at random were independently checked by a second coder. This second coder was naive as to the design and purpose of the research, and completed a training session including instruction in the coding conventions used for unclear frames, and supervised coding of one participant. Overall the agreement rate between the first and second coder was 98.03% (SD 1.56%), with the lowest

agreement rate on any individual participant being 94.83%. This was taken as evidence that hand coding was reliable.

Results

To provide an overview of the eye-movement data, the results will first be presented in the form of descriptive graphs followed by more detailed statistical analyses, separately for reflexives and pronouns. The response accuracy rates to the comprehension questions were high (children: mean 94.8%, SD 3.69%; adults: mean 97.0%, SD 3.04%) indicating that participants were attending to the auditory stimulus materials. For the eye-movement data, visual inspection of the proportion of looks to each of the five regions (the two animate characters, the inanimate object, the distracter, or the centre) during the first two seconds following the onset of the reflexive or pronoun showed that the pattern of looks to the inanimate object, the distracter and the centre were similar for both adults and children in both conditions, with few looks to the distracter and centre, and a high proportion of looks to the inanimate object, e.g., to 'popcorn' for trial (3). Over the two-second time window and across conditions, looks to the inanimate object accounted for 48% of all fixations. This is due to the fact that the inanimate object was the last entity mentioned before participants heard the reflexive or pronoun. In the following, only looks to the two potential antecedents will be further described and analysed.

Reflexives. Figures 1 and 2 show adults' and children's fixations of the two potential antecedents in the two experimental conditions (double-match/single-match) for the two seconds following the onset of the critical reflexive pronoun. The x-axis

displays the time in milliseconds from the onset of the reflexive for a 2000ms time window. The y-axis depicts the proportions of looks (subject means) to the two animate characters, i.e., the number of trials in which a participant fixated on a particular picture for each 40ms video frame as a proportion of the total number of trials in which they were looking at the screen. Note that because it takes approximately 200ms to program an eye-movement (Rayner, Slowiaczek, Clifton & Bertera, 1983), only changes in proportions of looks after 200ms can be attributed to participants hearing the reflexive.

//INSERT FIGURES 1 AND 2 ABOUT HERE//

For adults (Figure 1), the proportion of looks to the accessible antecedent rose in response to hearing the reflexive from around 200ms in the single-match and 500ms in the double-match condition, peaking around 800-1200ms. The adults' looks to the inaccessible antecedent remained stable throughout the entire time window, with only a brief rise in looks to the gender matching inaccessible antecedent in the double match condition between 200-500ms. The children (Figure 2) also showed an increase in looks to the accessible antecedent in response to hearing the reflexive, with a sharp increase between 400-700ms in the single-match condition and a more steady rise between 400-1000ms in the double-match condition. Children's looks to the inaccessible antecedent varied considerably between conditions. In the single-match condition in which the two potential antecedents differed in gender, proportions of looks to the inaccessible antecedent fell from around 400-800ms and remained low (< 10%). In contrast, in the double-match condition children's looks to the gender matching inaccessible antecedent rose in tandem with (and even exceeding) looks to

the accessible antecedent during the first 700ms before levelling off, remaining notably higher than in the single match condition for the remainder of the time window. Interference from the gender matching competitor antecedent on the children's fixation patterns was much stronger and more long-lasting than in the adult group, where there was only a slight sign of interference for a short period of time between 200-500ms. These differences suggest that children had more difficulty than adults in ruling out the inaccessible antecedent as a potential referent when gender cues were not informative.

To analyse the eye-movement data statistically, mixed-effects logistic regression models were applied using the 'R' software package, version 2.10.1 (R Development Core Team, 2010). Analyses were carried out on the raw data with no aggregation over time, conditions, participants or items (Baayen, Davidson & Bates, 2008), focusing on looks to the two potential animate antecedents between 200-2000ms from the onset of the reflexive. Following standard procedures, predictor variables were grand-mean centred, to avoid issues of collinearity between predictors. The binary dependent variable encoded whether the picture of one of the two antecedents was, or was not, fixated for each of the 40ms frames. Models were fit to test for subject and item random intercepts and random slopes for each factor as well as for experimental factors. An empty model was first fit to the data, then further predictors were added and potential improvements to the fit of model were tested. As the graphs clearly show that the relationship between looks and time is not linear, first, second and third order polynomials of time were tested as predictors to determine the model that best captured the pattern of changes over time. First a linear model was fitted (first order), then a second order (quadratic) polynomial of time was added to the model to produce a parabola (single curve). A third order (cubic)

polynomial of time was then added, producing an 'S' shaped curve (two curves), and the models were compared to assess which gave the best fit. Only predictors that led to a significant improvement in the fit of the model were retained, such that the best fit model was achieved. Models were first fit to the full data set (from both adults and children), then any interaction terms were further explored by analysing data from each group separately. Models fitted to data from the child group also tested whether age was a significant predictor of online performance.

//INSERT TABLE 2 HERE//

Table 2 shows the fixed effects from the best fitting mixed-effects logistic regression model fitted to the full data set (both adults and children). The negative coefficient for the significant main effect of Antecedent Type (= 'Ant(Inaccessible)') reflects the fact that participants fixated on the accessible antecedent more than the inaccessible antecedent, and the three-way interaction between Antecedent Type (accessible vs. inaccessible), Condition (single-match vs. double-match) and Group (adults vs. children) confirms that adults and children differed with regard to the extent to which they were distracted by a gender matching but structurally inaccessible competitor antecedent.

Before investigating these child/adult differences further, two additional analyses were performed to explore potential correlations among the measures. Firstly, because the data were analysed every 40ms, the data from a given sample might be (positively) correlated with those from immediately adjacent samples. To test for the local non-independence of the 40ms samples, an additional (more coarse-grained) analysis was performed in which the data were sampled every 200ms,

instead of every 40ms. Secondly, because increased looks to one of the two antecedents may result in decreased looks to the other character on the screen, the proportions of looks to the two antecedents might be (negatively) correlated with each other. Regarding the potential non-independence of looks to the two antecedents, it should be noted, however, that for the original analysis in Table 2, proportions of looks were calculated out of all of the looks to the 5 screen areas (4 pictures and centre), which means that the relationship between looks to the two antecedents is indirect, and a change in looks to one antecedent does not necessarily affect the looks to the other. Nevertheless, to examine this concern more directly, we performed an additional analysis on looks to the inaccessible antecedent only across the different conditions, i.e. on a subset of the data set, without using Antecedent Type as a factor. The new analysis (for 200ms samples) yielded the same three-way interaction between Antecedent Type, Condition and Group as the original (40ms) analysis (β = 0.731, St. Error = 0.193, z = 3.779, $p < .001$). Furthermore, even when only looks to the inaccessible antecedent were included, the Condition x Group interaction was still maintained (β = -0.992, St. Error = 0.424, z = -2.339, p = .019). Taken together, the results from these two additional analyses confirm the robustness of the effects.

In the following, we further examined the three-way interaction obtained in the original analysis (see Table 2). To explore this interaction, separate models were fitted to the data from the two participant groups.

//INSERT TABLE 3 HERE//

For the adult group, Table 3 shows a significant main effect of Antecedent Type, but no interaction with Condition, reflecting the fact that the adults looked

significantly more at the accessible antecedent than at the inaccessible one, irrespective of the inaccessible antecedent's gender. Thus, the numerical trend observed in Figure 1 of increased proportions of looks to the inaccessible antecedent between 200-500ms in the double-match condition did not turn out to be statistically reliable.

The child group also looked significantly more at the accessible antecedent than at the inaccessible one, but this pattern was modulated by Condition, with the positive coefficient for the significant Antecedent Type x Condition interaction showing that the contrast between looks to the two antecedent types was less (i.e., the negative slope was adjusted positively) in the double-match condition than the single match condition, confirming the numerical patterns seen in Figure 2.

Furthermore the negative coefficient for the interaction between Antecedent Type x Condition x Age shows that for older children this positive adjustment was smaller, as older children's pattern of looks across the two potential antecedents was less affected by the gender of the inaccessible antecedent than younger children's. For illustrative purposes, the proportions of looks to the accessible and inaccessible antecedents in the double-match condition are plotted separately for our six-to-seven-year-old (Year 2) and eight-to-nine year-old (Year 4) child participants in Figures 3a and 3b.

//INSERT FIGURE 3 ABOUT HERE//

Figure 3 shows that whilst both the younger and the older children experienced competition between the two antecedents during the first 700-800ms after hearing the reflexive, the older children's proportion of looks to the inaccessible

antecedent started levelling off about 200ms earlier. The younger children also appeared to be more strongly distracted by the inaccessible antecedent again during later (> 1100ms) time windows.

In summary, our results show that children's online referential decisions were influenced by an inaccessible but highly prominent competitor antecedent (viz., the main clause subject) if it matched in gender with the reflexive. Furthermore, the effect of the inaccessible competitor was found to significantly decline as the children grew older.

Pronouns. Figures 4 and 5 present the adults' and children's descriptive results for the pronouns, showing which potential antecedent participants fixated upon on hearing the pronoun in the two experimental conditions (double-match/single-match). Note that, unlike for reflexives, the accessible antecedent here is the main clause subject (i.e., the pronoun *he/she*), which refers back to the subject of the lead-in sentence (e.g., *Peter/Susan* in (3)).

//INSERT FIGURES 4 AND 5 ABOUT HERE//

Figures 4 and 5 show that the referent of the main clause subject (the accessible antecedent) attracted a considerably higher proportion of looks than the embedded subject (the latter being structurally inaccessible, according to Principle B), which represents the opposite of the pattern seen for reflexives, where the picture depicting the embedded subject was fixated more often than the one showing the main clause subject (the latter being ruled out by Principle A). These contrasts are parallel for children and adults. Figures 3 and 4 also show a similar time course of looks to the

accessible and the inaccessible antecedents for children and adults, with the proportion of looks to the accessible antecedent gradually increasing from 200ms, and peaking from 600ms to 1200ms, in both participant groups.

These observations suggest that both adults' and children's gaze patterns were influenced by the different binding properties of reflexives and pronouns. At the same time, the graphs for both participant groups also show some influence of the inaccessible antecedent, indicated by differing patterns of looks across the two conditions (single-match/double-match). Specifically, adults showed fewer looks to the accessible antecedent, and both participant groups showed more looks to the inaccessible one, in the double-match than in the single-match condition. Adults' proportions of looks to the inaccessible antecedent remained low and approximately stable over the whole 2000ms time window, whereas children's looks to the inaccessible antecedent *increased* from 200-800ms before levelling off. These differences suggest that the gender of the inaccessible referent affected both adults' and children's looks, and also that adults and children responded differently to the two conditions. The data on pronouns were statistically analysed in the same way as those on reflexives.

//INSERT TABLE 4 HERE//

Table 4 shows the fixed effects from the best fitting mixed-effects logistic regression model fitted to the full data set from both adults and children. The results revealed a significant main effect of Antecedent Type, reflecting the fact that participants fixated on the accessible antecedent more than the inaccessible antecedent and, more importantly, a three-way interaction between Antecedent Type, Condition

and Group, showing that adults and children differ with regard to the way in which they were distracted by a gender matching but binding inaccessible antecedent. To further explore this interaction, separate models were fitted to the data from the two participant groups.

//INSERT TABLE 5 HERE//

Results for both children and adults show a significant main effect of Antecedent Type and a two-way interaction between Antecedent Type and Condition, confirming that the pattern of looks to the two antecedents was modulated by the gender of the (inaccessible) competitor referent, with greater differences between looks to the two potential antecedents in the single-match compared to the double-match condition.[1]

Furthermore, the interaction with 'Age' shows that older children were less distracted by the gender matching but inaccessible antecedent than younger children. Figure 6 illustrates that only the younger children showed an almost equally steep rise in their proportions of looks to both antecedents during the first 800ms after encountering the pronoun, and compared to the older children, the younger ones showed a relatively smaller advantage for the accessible antecedent throughout the initial 1400ms. The older children, in contrast, showed a consistently low proportion of looks to the inaccessible antecedent, similar to adults.

[1] The interaction with Group reported in Table 4 reflects child/adult differences in the proportions of looks to the *accessible* antecedent, with the adults – but not the children – fixating the accessible antecedent more in the single-match compared to the double-match condition (see Figure 4 vs. Figure 5).

//INSERT FIGURE 6 ABOUT HERE//

Finally, recall that some children occasionally made errors in the offline task in that they incorrectly accepted local coreference interpretations for pronouns. To assess how this affected children's eye-movement data in the online experiment, we performed the same analysis as before, except that instead of 'Age' the offline accuracy scores (for the pronoun/double-match/'*no*' response) were entered into the regression model. This analysis yielded a three-way interaction between Antecedent Type, Condition and Offline Score (β = -1.280, St. Error = 0.329, z = -3.885, p < .001), indicating that children who achieved lower offline scores were also more distracted on-line by a gender matching competitor referent.

In summary, we found that both adults' and children's online interpretation of pronouns was significantly affected by an inaccessible but gender-matching competitor antecedent, with younger children experiencing comparatively more interference than older ones.

General Discussion

The current study sought to gather new experimental evidence on children's referential interpretations for reflexives in comparison to pronouns during real-time sentence comprehension. Examining groups of six-to-nine-year-old children and adult controls, we specifically asked how and when an inaccessible competitor antecedent influences anaphor resolution. Our most interesting finding was a contrast between children's ultimate interpretations of reflexives and their online processing patterns. Whilst children's offline performance on reflexives in experiment 1 was adult-like,

the online data from experiment 2 showed that they were temporarily more distracted than adults by a gender-matching but structurally inaccessible antecedent.

Experiment 1 furthermore showed that for non-reflexive pronouns, children sometimes accepted (incorrect) coreference interpretations with a gender-matching local antecedent. The results from experiment 2 revealed that during online processing, both adults and children experienced significant interference from a structurally inaccessible antecedent where gender cues failed to disambiguate the pronoun.

In what follows, we will discuss these findings with respect to three wider issues, (i) what types of cues children (and adults) rely on for their referential interpretations of reflexives and pronouns, (ii) when during processing these different types of cues are used, and (iii) how the observed developmental changes from child to adult can be explained.

Binding Principles and Discourse Prominence

Our results show that both adults' and children's offline and online performance on reflexives and pronouns was sensitive to structurally defined constraints on coreference, notably Binding Principles A and B. Neither discourse-level nor surface cues such as the relative distance between the anaphoric element and its potential antecedents can, by themselves, explain the observed pattern of results. If, for example, participants were simply favouring the linearly closest potential antecedent, this would have yielded correct results for reflexives in experiment 2, but not for pronouns. Likewise, if they preferred coreference with the most prominent referent in the discourse, i.e., the matrix subject of both the lead-in and the critical sentences, this would account for the results on pronouns, but not for those on

reflexives. Instead, the observed patterns of preferences can better be explained in terms of binding constraints, in that structurally accessible antecedents were fixated more than inaccessible ones, for both reflexives and pronouns. More generally, our results confirm that binding principles do not only constrain ultimate referential interpretations but also guide the online processing of reflexives and pronouns in both adults and children (e.g., Nicol & Swinney, 1989; McKee et al., 1993; Sturt, 2003; Love et al., 2009; Xiang et al., 2009).

Previous experimental research with adults has found that online referential interpretations for reflexives and pronouns are not only influenced by binding constraints but also by a number of other (morphological, semantic, and discourse-level) properties of the linguistic environment, which affect a potential antecedent's relative 'accessibility' (e.g., Badecker & Straub, 2002; Sturt, 2003; Koornneef, 2008; Cunnings & Felser, in press). Our results provide further support for this finding. For pronouns, both adults and children fixated the inaccessible antecedent more in the double-match than in the single-match condition, suggesting that they were less confident in their referential interpretations when both antecedents matched in gender with the pronoun. Similarly, for reflexives, an inaccessible competitor referent also caused distraction, particularly for children. In the child data, significantly more looks to the inaccessible antecedent were found when it matched in gender with the reflexive compared to when this was not the case. Although this competition was less pronounced in the adult data, it was still visible as a numerical trend from 200-600ms (compare Figure 1).

Inaccessible antecedents were promoted as potential referents for pronouns or reflexives by a number of factors, most notably gender (in the double-match conditions), recency (for pronouns) and prominence (for reflexives). In the case of

pronouns, the structurally inaccessible antecedent was the embedded subject, the linearly closest potential antecedent. Although ruled out by Principle B, interpreting a pronoun as coreferential with the embedded subject was thus likely to be the most memory-friendly option. For reflexives, on the other hand, the inaccessible antecedent was favoured by virtue of being the most prominent referent in the linguistic context, in that it was the first-mentioned referent, was referred to twice prior to the occurrence of the reflexive, and was also located in subject position. Note, however, that even though recency and a potential antecedent's relative discourse-prominence affected the referential interpretation of both pronouns and reflexives causing temporary distraction, they did not override the effects of binding constraints, as witnessed by the fact that inaccessible antecedents consistently attracted fewer looks than accessible ones for both types of anaphor.

Time-Course Issues

One hypothesis concerning the temporal dynamics of referential interpretations for pronouns and reflexives was that binding principles should be applied early during processing and act as an initial filter, ruling out potential referents that are not structurally accessible. This hypothesis was originally proposed by Nicol and Swinney (1989) and backed up by findings from cross-modal priming experiments with adults. Similar experiments were carried out with children by McKee et al. (1993) and Love et al. (2009), leading to the same conclusion - that binding principles are applied early during processing. Note, however, that these studies only examined priming effects from one of the two potential antecedents that were present in the context, so could not tell us anything about the extent to which

listeners might have experienced competition from an alternative referent when encountering the reflexive or pronoun.

The present set of findings seems to provide little support for the binding-as-initial-filter hypothesis. Both the child and the adult data indicated competition from the inaccessible antecedent for the interpretation of pronouns, which was visible from 200ms onwards, i.e., the earliest point at which eye-movements reflect processing of the pronoun (compare Figures 4 and 5). For reflexives, evidence of distraction caused by a gender-matching but inaccessible antecedent was more obvious in the child than the adult data. Again, this evidence was seen very early, from 200ms for adults (Figure 1) and from 400ms for children (Figure 2). There was no evidence that effects of the non-local competitor antecedent were considerably delayed in the children relative to the adult controls, as was the case for the ambiguous 'short-distance' pronouns examined in Sekerina et al.'s (2004) study. Thus, according to our eyemovement results, binding principles did *not* function as an initial filter on potential referential interpretations.

One potential explanation for the early competition our participants experienced from the inaccessible antecedent could be that looks to a gender-matched inaccessible antecedent were initiated as soon as participants had heard the initial syllable *him* or *her*, that is, before they realized that they were hearing a reflexive rather than a pronoun. However, this explanation is unlikely because a phonetic analysis of our auditory materials revealed that the first syllable of a reflexive is very different from a non-reflexive pronoun in fluent speech, with the vowel in the former reduced to a brief 'schwa'. A comparison of the segments *him*/*her* (within reflexive pronouns) to the words *him*/*her* (non-reflexive pronouns) in the (double-match conditions of the) spoken materials of experiment 2 revealed significant differences

both in terms of their duration (in ms) and their intensity (in dB), with the *him*/*her* (reflexive) segments being significantly reduced in duration and intensity compared to the *him*/*her* pronouns (158 vs. 275 ms, $t(23) = 20.34$, $p < .001$, 69.6 vs. 71.1db, $t(23) = 3.38$, $p = .003$). Given these differences, participants are unlikely to mistake the first syllable of the reflexive for the full pronoun *him* or *her*.

On the other hand, it was also *not* the case, either for children or for adults, that sensitivity to binding principles was delayed relative to the point in time at which properties of the inaccessible antecedent were considered. Instead, our data provide support for the view (e.g., Badecker & Straub, 2002) that from the earliest measurable point in time referential interpretations are determined by binding constraints in tandem with other sources of information.

Developmental Issues

Developmental changes from child to adult were seen in the data of the present study for both reflexives and pronouns. The most striking finding was the contrast between the results of the offline and the online experiments for reflexives, with the children performing adult-like in the offline but not in the online task. Developmental differences were also seen in the results for pronouns, with children sometimes violating Principle B in the offline task, and younger children performing less adult-like compared to older children in the online task.

With respect to reflexives, our finding that six-to-nine-year olds obeyed Principle A and showed adult-like performance in the offline task is not particularly surprising, as many previous studies have also reported correct performance on reflexives even for three-year-old children (e.g., Guasti, 2002). Clearly, children's ultimate referential interpretations for reflexives are determined by Principle A.

Online reflexive anaphor resolution, in contrast, does not yet appear to be fully adult-like in six-to-nine-year-old children. We found the children – but not the adult controls - to be significantly distracted by an inaccessible (i.e., non-local) competitor antecedent that matched in gender with the reflexive. Recall that this antecedent was highly prominent in the linguistic environment in that it was the first-mentioned character, was located in subject position, and was referred to twice prior to the occurrence of the reflexive. There is evidence from previous studies that children's online referential interpretations for pronouns are affected by discourse prominence (e.g., Song & Fisher, 2005, 2007), and our results indicate that this is also true of reflexives. In cases in which gender cues are uninformative, children are more tempted than adults to interpret the reflexive as coreferential with the most highly activated potential antecedent in the discourse. The application of Binding Principle A, on the other hand, requires that the highly activated non-local antecedent be deactivated in favour of the accessible (but less prominent) local antecedent. Children seem to struggle with this deactivation process, resulting in the kind of interference from the inaccessible antecedent that we saw in the online experiment. This interpretation is consistent with previous studies suggesting that children have more difficulty than adults weighing up and integrating information from different sources and revising initial parsing decisions (e.g., Choi & Trueswell, 2010, Trueswell et al., 1999, Traxler, 2002). Consequently, children are more distracted than adults in cases in which Binding Principles and discourse prominence provide conflicting cues.

The younger children's relatively greater difficulty inhibiting inaccessible antecedents during processing might be due to developmental changes in executive function or cognitive control (Novick, Trueswell, & Thompson-Schill, 2005; see Mazuka, Jincho, & Oisho, 2009, and Novick, Trueswell, & Thompson-Schill, 2010,

for reviews). Also using eye-tracking during listening, Choi and Trueswell (2010), for example, report that four-to-five year-old Korean-speaking children, unlike adults, had difficulty recovering from initial misinterpretations of 'garden-path' sentences. In the light of these and similar earlier findings from English (e.g., Trueswell et al, 1999), the authors argue that children's limited cognitive control abilities may prevent them from inhibiting misinterpretations even in the face of clear disambiguating evidence.

In short, the comparison of children' and adults' performance on reflexives revealed that although children demonstrated more difficulties than adults during online anaphor resolution, they must have been able to resolve the conflicting information at a later stage and arrive at the same ultimate interpretation as adults, in accordance with Principle A. Furthermore, the impact of the competitor antecedent on children's online performance for reflexives significantly decreased with age, with the older children in the 6-9 year-old age range showing fewer looks to a gender-matched inaccessible antecedent than the younger children. Finally, recall that the adult group's eye-movement patterns also showed signs of early temporary distraction from the inaccessible antecedent, albeit to a much lesser degree than the children and not significantly so. From these observations, we conclude that what distinguishes children and adults in this domain is children's ability to handle and prioritize competing sources of information, or to inhibit misinterpretations, which seems to initially lag behind, and to develop throughout the primary school years to adult level.

As regards pronouns, our results suggest that both children and adults experienced measurable interference from a gender-matching (but structurally inaccessible) local competitor antecedent during processing. The degree to which children were distracted by the inaccessible antecedent in experiment 2 decreased

with age, but the results from our offline task (experiment 1) indicate that not all children were able to ultimately rule out coreference between a pronoun and a local competitor antecedent (i.e., apply Principle B). Unlike Love et al. (2009), who examined children from a wider age range than we did, we found no evidence in our data to suggest that six-to-nine-year-old children's offline mastery of Principle B was significantly correlated with age. Our eye-movement data suggest that even adults had more difficulty eliminating an inaccessible competitor antecedent from the candidate set for pronouns compared to reflexives, although unlike for the children, this difficulty was not reflected in their offline antecedent choices. This is what we might expect given that unlike Binding Principle A, Principle B fails to identify a unique referent for a pronoun, whose interpretation requires recourse to, and integration of, additional information sources. Furthermore, experimental evidence indicates that the interpretation of pronouns is more strongly affected by non-syntactic factors compared to reflexives. Also using eye-tracking during listening, Kaiser et al. (2009), for example, found that adults' on-line interpretation of reflexives and pronouns in picture noun phrases was influenced both by syntactic constraints and thematic role information, even if the syntactically appropriate antecedent was ultimately chosen, and that the influence of thematic role based constraints was relatively stronger for pronouns than for reflexives. Taking into account both Kaiser et al.'s and our findings, it is conceivable that listeners generally have more difficulty inhibiting a semantically or pragmatically prominent but syntactically inappropriate antecedent for pronouns compared to reflexives (whose interpretation is primarily determined by syntactic constraints) during online processing.

In summary, our results demonstrate that children had more difficulty than adults deactivating a highly prominent, gender-matching competitor antecedent when

interpreting reflexives, which is in line with previous findings suggesting that children have more difficulty than adults controlling multiple information sources. For pronouns, both children and adults showed reliable evidence of temporary competition between two gender-matching antecedents in their eye-movement patterns, with not all of the children able to ultimately reject a structurally inaccessible antecedent when interpreting pronouns offline.

Conclusion

We set out to determine how binding constraints interact with other factors (notably discourse prominence) during children's and adults' processing of reflexives in comparison to pronouns. Our eye-movement results indicate that unlike what several studies have reported for monolingual adults, Binding Principle A does not act as an initial filter on children's interpretation of reflexives but interacts with discourse-level information from early on during processing. Evidence of antecedent competition was also seen for pronouns in both the children's and adults' eye-movement records, suggesting that structurally inaccessible antecedents are not immediately filtered out by Principle B, either. We also found age effects indicating that younger children had more difficulty than older ones homing in on the structurally accessible antecedent, both for reflexives and pronouns, which might be attributable to the relatively late development of relevant cognitive control abilities. Our findings show that children who behave adult-like in offline tasks do not necessarily process anaphors in an adult-like way, thus underscoring the need for gathering online data to gain a more comprehensive picture of children's mastery of structural constraints on interpretation, and their ability to access and integrate different information sources during comprehension.

Acknowledgments

This research was funded by an ESRC postgraduate studentship to KC and a University of Essex Research Promotion Fund grant to CF and HC. We thank children and staff of Colville Primary School, St. Matthew's Primary School, King's Hedges School and Dame Bradbury's school for taking part in this research. We are also grateful to three anonymous *JML* reviewers and the members of the Psycholinguistics Research Group at the University of Essex for helpful comments and feedback.

References

Arnold, J., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes, 22*(4), 527-565.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412.

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*(4), 748-769.

Brugman, H., & Russel, A. (2004). *Annotating multimedia/multi-modal resources with ELAN.* Paper presented at the Fourth International Conference on Language Resources and Evaluation.

Chien, Y., & Wexler, K. (1990). Children's Knowledge of Locality Conditions in Binding as Evidence for the Modularity of Syntax and Pragmatics. *Language Acquisition, 1*(3), 225-295.

Choi, Y., & J. C. Trueswell (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology, 106*(1), 41-61.

Chomsky, N. (1981). *Lectures in Government and Binding*: Dordrecht: Foris.

Cunnings, I. & C. Felser (in press). The role of working memory in the processing of reflexives. *Language and Cognitive Processes.*

Felser, C., & Clahsen, H. (2009). Grammatical processing of spoken language in child and adult language learners. *Journal of Psycholinguistic Research, 38*(3), 305-319.

Fernald, A., Zangl, R., Portillo, A., & Marchman, V. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. *Language Acquisition and Language Disorders, 44*, 97.

Forster, K., & Forster, J. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116.

Guasti, M. (2002). *Language acquisition: The Growth of Grammar*: MIT Press.

Harris, T., Wexler, K., & Holcomb, P. (2000). An ERP investigation of binding and coreference. *Brain and Language, 75*(3), 313-346.

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition, 112*, 55-80.

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

Love, T., Walenski, M., & Swinney, D. (2009). Slowed speech input has a differential impact on on-line and off-line processing in children's comprehension of pronouns. *Journal of Psycholinguistic Research, 38*(3), 285-304.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2009). Pronoun coreferencing errors: Challenges for generativist and usage-based accounts. *Cognitive Linguistics, 20*(3), 599-626.

Mazuka, R., Jincho, N., & Oishi, H. (2009). Development of executive control and language processing. *Language and Linguistics Compass, 3*(1), 59-89.

McKee, C., Nicol, J., & McDaniel, D. (1993). Children's application of binding during sentence processing. *Language and Cognitive Processes, 8*(3), 265-290.

Nicol, J. (1988). *Coreference processing during sentence comprehension*: PhD Thesis. Massachusetts Institute of Technology, Dept. of Brain and Cognitive Sciences.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*(1), 5-19.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 5*(3), 263-281.

Novick, J. M., Trueswell, J. C., Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass 4*(10), 906-924.

Pyykkonen, P., Matthews, D., & Järvikivi, J. (2010). Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution. *Language and Cognitive Processes, 25*(1), 115-129.

R Development Core Team (2010). R: A language and environment for statistical computing, *R Foundation for Statistical Computing, Vienna, Austria.* (pp. URL http://www.R-project.org).

Rayner, K., Slowiaczek, M. L., Clifton, C., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance, 9*(6), 912-922.

Reinhart, T. (2006). *Interface strategies*: MIT Press.

Sekerina, I., Stromswold, K., & Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns? *Journal of child language, 31*(01), 123-152.

Song, H., & Fisher, C. (2005). Who's "she"? - Discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language, 52*(1), 29-57.

Song, H., & Fisher, C. (2007). Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua, 117*(11), 1959-1987.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48*(3), 542-562.

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., et al. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers, 35*(4), 621.

Thornton, R., & Wexler, K. (1999). *Principle B, VP ellipsis, and interpretation in child grammar*: The MIT Press.

Traxler, M. (2002). Plausibility and subcategorization preference in children's processing of temporarily ambiguous sentences: Evidence from self-paced reading. *The Quarterly Journal of Experimental Psychology 55A*(1), 75-96.

Trueswell, J. (2008). Using eye movements as a developmental measure within psycholinguistics. *Developmental Psycholinguistics: On-line methods in children's language processing*, 73-96.

Trueswell, J., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*(2), 89-134.

Van der Lely, H., & Stollwerck, L. (1997). Binding theory and grammatical specific language impairment in children. *Cognition, 62*(3), 245-290.

Van Rij, J., Van Rijn, H., & Hendriks, P. (2010). Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language, 37*(03), 731-766.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language, 108*(1), 40-55.

Appendix A: Auditory materials for experiment 2

The experimental trials for double-match condition with reflexives are shown below. As explained in the text, the same materials were used with pronouns, and with different characters, for the single-match condition.

(1) Peter was waiting outside the corner shop. He watched as Mr. Jones bought a huge box of popcorn for himself over the counter

(2) Peter was laying the dinner table. He asked whether Mr. Jones could fetch a large clean plate for himself from the other room

(3) Peter was spending a day at the beach. He was amazed to see that Mr. Jones had built a magnificent castle for himself from just sand and water

(4) Peter was visiting a dairy farm. He watched as Mr. Jones made some special rich cheese himself from very fresh goat's milk

(5) Peter was very interested in water sports. He wondered whether Mr. Jones had bought the new canoe for himself that had been shown on TV

(6) Peter was feeling a little peckish. He saw that Mr. Jones was cracking a huge walnut for himself with a rusty nutcracker

(7) Susan was sitting at the kitchen table. She watched as Mrs. White made a large salami pizza for herself as a special treat

(8) Susan was sitting by the swimming pool. She noticed that Mrs. White had bought a large ice cream cone for herself at the hotel reception

(9) Susan was feeling very hungry. She watched as Mrs. White slowly peeled a large juicy pear for herself with a sharp kitchen knife

(10) Susan watched the snow falling outside. She could hear that Mrs. White was making a log fire for herself in the master bedroom

(11) Susan was watching the customers in the post office. She wondered why Mrs. White was sending a small package to herself that Friday afternoon

(12) Susan was standing in the kitchen. She noticed that Mrs. White was boiling a large egg for herself on the brand new hob

(13) Mr. Jones was listening very hard. He knew that Peter was playing some classical music to himself on the new piano

(14) Mr. Jones was looking forward to Halloween. He knew that Peter had carved out a giant pumpkin for himself that would make a nice lantern

(15) Mr. Jones was very nervous before the TV interview. He suddenly realized that Peter was attaching a microphone to himself and that a light was flashing

(16) Mr. Jones knew that winter wasn't far away. He smiled when Peter started knitting a long woollen scarf for himself by the old fireplace

(17) Mr. Jones was very tired. He barely noticed that Peter was noisily running a bath for himself shortly after dinner

(18) Mr. Jones was fascinated by technology. He was impressed when Peter managed to build a toy robot for himself from scrap metal and cardboard

(19) Mrs. White was looking around the toy shop. She frowned when |Susan bought a brightly coloured kite for herself that cost a small fortune

(20) Mrs. White used to teach geography. She was pleased to see that Susan had drawn a large and colourful map for herself that would look nice on the wall

(21) Mrs. White is a law-abiding citizen. She was shocked when Susan stole an expensive bottle of wine for herself from the supermarket

(22) Mrs. White was making some lemonade. She asked whether Susan could go and fetch a clean glass for herself from the kitchen cupboard

(23) Mrs. White couldn't stand the sight of blood. She was relieved when Susan went to get a large plaster for herself that covered the small scratch

(24) Mrs. White is a fashion designer. She was pleased to see that Susan had made a beautiful belt for herself from soft brown leather

(25) Peter was waiting outside the corner shop. He watched as Mr. Jones bought a huge box of popcorn for himself over the counter

(26) Peter was laying the dinner table. He asked whether Mr. Jones could fetch a large clean plate for himself from the other room

(27) Peter was spending a day at the beach. He was amazed to see that Mr. Jones had built a magnificent castle for himself from just sand and water

(28) Peter was visiting a dairy farm. He watched as Mr. Jones made some special rich cheese himself from very fresh goat's milk

(29) Peter was very interested in water sports. He wondered whether Mr. Jones had bought the new canoe for himself that had been shown on TV

(30) Peter was feeling a little peckish. He saw that Mr. Jones was cracking a huge walnut for himself with a rusty nutcracker

(31) Susan was sitting at the kitchen table. She watched as Mrs. White made a large salami pizza for herself as a special treat

(32) Susan was sitting by the swimming pool. She noticed that Mrs. White had bought a large ice cream cone for herself at the hotel reception

(33) Susan was feeling very hungry. She watched as Mrs. White slowly peeled a large juicy pear for herself with a sharp kitchen knife

(34) Susan watched the snow falling outside. She could hear that Mrs. White was making a log fire for herself in the master bedroom

(35) Susan was watching the customers in the post office. She wondered why Mrs. White was sending a small package to herself that Friday afternoon

(36) Susan was standing in the kitchen. She noticed that Mrs. White was boiling a large egg for herself on the brand new hob

(37) Mr. Jones was listening very hard. He knew that Peter was playing some classical music to himself on the new piano

(38) Mr. Jones was looking forward to Halloween. He knew that Peter had carved out a giant pumpkin for himself that would make a nice lantern

(39) Mr. Jones was very nervous before the TV interview. He suddenly realized that Peter was attaching a microphone to himself and that a light was flashing

(40) Mr. Jones knew that winter wasn't far away. He smiled when Peter started knitting a long woollen scarf for himself by the old fireplace

(41) Mr. Jones was very tired. He barely noticed that Peter was noisily running a bath for himself shortly after dinner

(42) Mr. Jones was fascinated by technology. He was impressed when Peter managed to build a toy robot for himself from scrap metal and cardboard

(43) Mrs. White was looking around the toy shop. She frowned when |Susan bought a brightly coloured kite for herself that cost a small fortune

(44) Mrs. White used to teach geography. She was pleased to see that Susan had drawn a large and colourful map for herself that would look nice on the wall

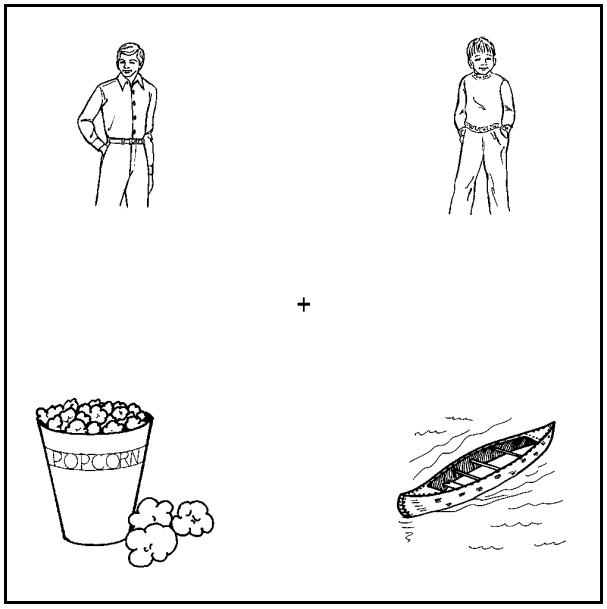
(45) Mrs. White is a law-abiding citizen. She was shocked when Susan stole an expensive bottle of wine for herself from the supermarket

(46) Mrs. White was making some lemonade. She asked whether Susan could go and fetch a clean glass for herself from the kitchen cupboard

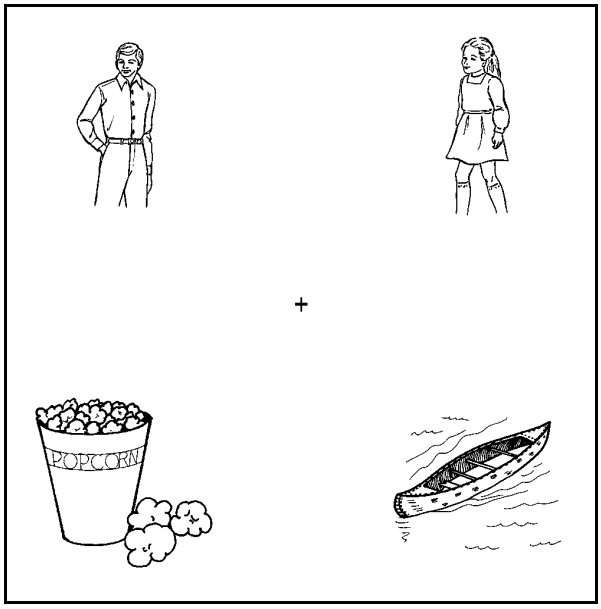
(47) Mrs. White couldn't stand the sight of blood. She was relieved when Susan went to get a large plaster for herself that covered the small scratch

(48) Mrs. White is a fashion designer. She was pleased to see that Susan had made a beautiful belt for herself from soft brown leather

Appendix B: Example visual displays.



Visual display for double-match condition



Visual display for single-match condition

Table 1: Mean percentages of correct (*yes* and *no*) responses in the child data in Experiment 1.

	Yes response		*No* response	
	Single-match	Double-match	Single-match	Double-match
Reflexives	99.50 (3.16)	97.00 (7.23)	99.50 (3.16)	99.00 (4.41)
Pronouns	99.00 (4.41)	94.50 (14.31)	98.00 (6.08)	86.50 (18.89)

Table 2: Fixed-effects from best fitting mixed-effects logistic regression model fit to data from both adults and children for reflexives, Experiment 2.

Fixed Effects	Estimate	St.Error	z Value	p Value
(Intercept)	-1.215	0.170	-7.143	<0.001
Linear time	-0.153	0.146	-1.053	0.292
Quadratic time	-0.821	0.139	-5.890	<0.001
Cubic time	0.292	0.195	1.497	0.134
Ant(Inaccessible)	-0.935	0.199	-4.705	<0.001
Condition(Double-Match)	-0.063	0.140	-0.447	0.655
Group(Children)	-0.380	0.220	-1.731	0.083
Ant(Inaccessible) x Condition(Double-Match)	0.043	0.063	0.673	0.501
Ant(Inaccessible) x Group(Children)	-0.158	0.220	-0.715	0.475
Condition(Double-Match) x Group(Children)	-0.044	0.193	-0.228	0.819
Ant(Inaccessible) x Condition(Double-Match) x Group(Children)	0.758	0.091	8.346	<0.001

Formula in R: DepVar~ Linear time + Quadratic time + Cubic time + Ant * Condition * Group + (1 + Ant + Condition + Linear time + Quadratic time + Cubic time |Part) + (1 + Ant + Group + Linear time + Quadratic time + Cubic time |Item)

Table 3: Fixed-effects from best-fitting models fit to adult and child data separately for reflexives, Experiment 2.

Fixed Effects	Estimate	St.Error	z Value	p Value
Adults				
(Intercept)	-1.308	0.192	-6.829	<0.001
Linear time	-0.115	0.145	-0.796	0.426
Quadratic time	-0.815	0.173	-4.703	<0.001
Ant(Inaccessible)	-0.910	0.232	-3.917	<0.001
Condition(Double-Match)	-0.081	0.147	-0.548	0.584
Formula in R: DepVar~ Linear time + Quadratic time + Ant + (1 + Ant + Linear time + Quadratic time |Part) + (1 + Ant + Linear time + Quadratic time |Item)				
Children				
(Intercept)	-1.553	0.183	-8.485	<0.001
Linear time	-0.207	0.215	-0.966	0.334
Quadratic time	-0.950	0.224	-4.250	<0.001
Cubic time	0.631	0.319	1.981	0.048
Ant(Inaccessible)	-1.225	0.248	-4.936	<0.001
Condition(Double-Match)	-0.095	0.137	-0.690	0.490
Age	-0.036	0.147	-0.246	0.806
Ant(Inaccessible) x Condition(Double-Match)	0.830	0.066	12.535	<0.001
Ant(Inaccessible) x Age	0.045	0.127	0.356	0.722
Condition(Double-Match) x Age	0.022	0.124	0.179	0.858
Ant(Inaccessible) x Condition(Double-Match) x Age	-0.154	0.070	-2.205	0.027
Formula in R: DepVar~ Linear time + Quadratic time + Cubic time + Ant * Condition * Age + (1 + Ant + Condition + Linear time + Quadratic time + Cubic time |Part) + (1 + Ant + Age + Linear time + Quadratic time + Cubic time |Item)				

Table 4: Fixed-effects from best fitting mixed-effects logistic regression model fit to data from both adults and children for pronouns, Experiment 2.

Fixed Effects	Estimate	St.Error	z Value	p Value
(Intercept)	-1.719	0.279	-6.154	<0.001
Linear time	-0.053	0.144	-0.365	0.715
Quadratic time	-0.595	0.160	-3.727	<0.001
Cubic time	0.327	0.206	1.588	0.112
Ant(Inaccessible)	-0.822	0.246	-3.335	0.001
Condition(Double-Match)	-0.253	0.216	-1.172	0.241
Group(Children)	-0.308	0.325	-0.950	0.342
Ant(Inaccessible) x Condition(Double-Match)	0.540	0.069	7.811	<0.001
Ant(Inaccessible) x Group(Children)	0.246	0.249	0.987	0.324
Condition(Double-Match) x Group(Children)	0.220	0.302	0.727	0.467
Ant(Inaccessible) x Condition(Double-Match) x Group(Children)	-0.220	0.097	-2.273	0.023

Formula in R: DepVar~ Linear time + Quadratic time + Cubic time + Ant * Condition * Group + (1 + Ant + Condition + Linear time + Quadratic time + Cubic time |Part) + (1 + Ant + Group + Linear time + Quadratic time + Cubic time |Item)

Table 5: Fixed-effects from best-fitting models fit to adult and child data separately for pronouns, Experiment 2.

Fixed Effects	Estimate	St.Error	z Value	p Value
Adults				
(Intercept)	-1.883	0.304	-6.188	<0.001
Linear time	-0.080	0.201	-0.400	0.689
Quadratic time	-0.373	0.192	-1.938	0.053
Cubic time	0.478	0.249	1.923	0.054
Ant(Inaccessible)	-0.768	0.319	-2.410	0.016
Condition(Double-Match)	-0.260	0.249	-1.044	0.296
Ant(Inaccessible) x Condition(Double-Match)	0.508	0.071	7.167	<0.001
Formula in R: DepVar~ Linear time + Quadratic time + Cubic time + Ant * Condition + (1 + Ant + Condition + Linear time + Quadratic time + Cubic time |Part) + (1 + Ant + Linear time + Quadratic time + Cubic time |Item)				
Children				
(Intercept)	-1.987	0.287	-6.929	<0.001
Linear time	-0.111	0.263	-0.423	0.672
Quadratic time	-1.039	0.268	-3.880	<0.001
Cubic time	0.351	0.384	0.915	0.360
Ant(Inaccessible)	-0.583	0.235	-2.479	0.013
Condition(Double-Match)	0.043	0.202	0.212	0.832
Age	-0.696	0.218	-3.200	0.001
Ant(Inaccessible) x Condition(Double-Match)	0.238	0.069	3.458	0.001
Ant(Inaccessible) x Age	0.425	0.154	2.770	0.006
Condition(Double-Match) x Age	0.427	0.191	2.235	0.025
Ant(Inaccessible) x Condition(Double-Match) x Age	-0.359	0.076	-4.738	<0.001
Formula in R: DepVar~ Linear time + Quadratic time + Cubic time + Ant * Condition * Age + (1 + Ant + Condition + Linear time + Quadratic time + Cubic time |Part) + (1 + Ant + Age + Linear time + Quadratic time + Cubic time |Item)				

Figure 1: Adults' proportions of looks to potential antecedents on hearing the reflexive (subject means).

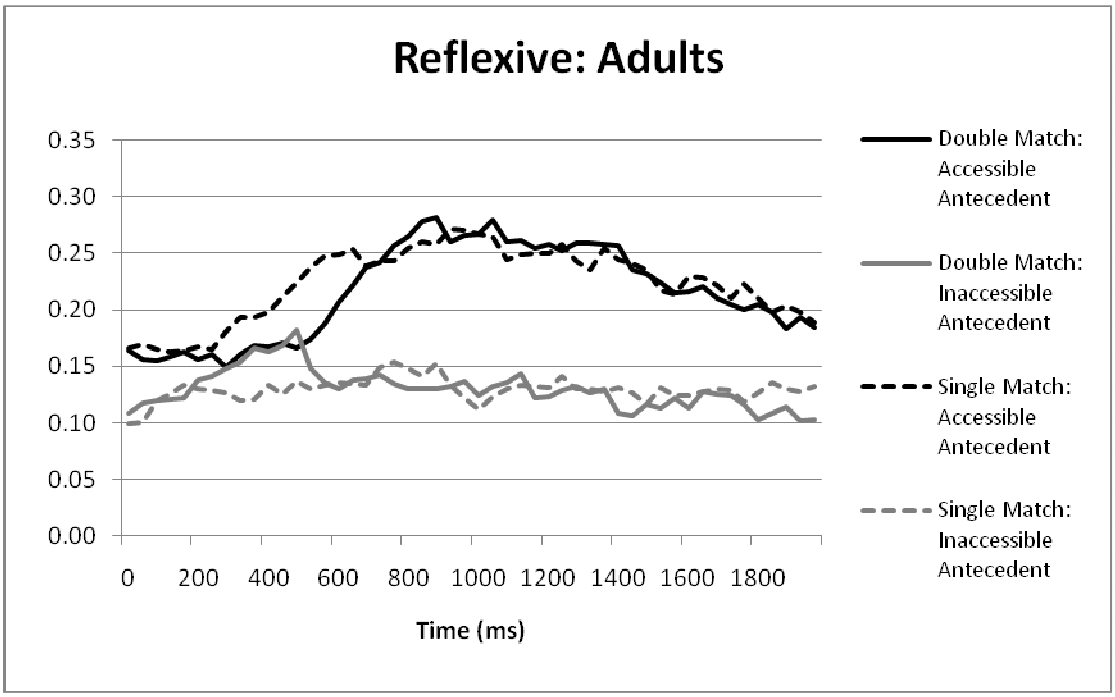


Figure 2: Children's proportions of looks to potential antecedents on hearing the reflexive (subject means).

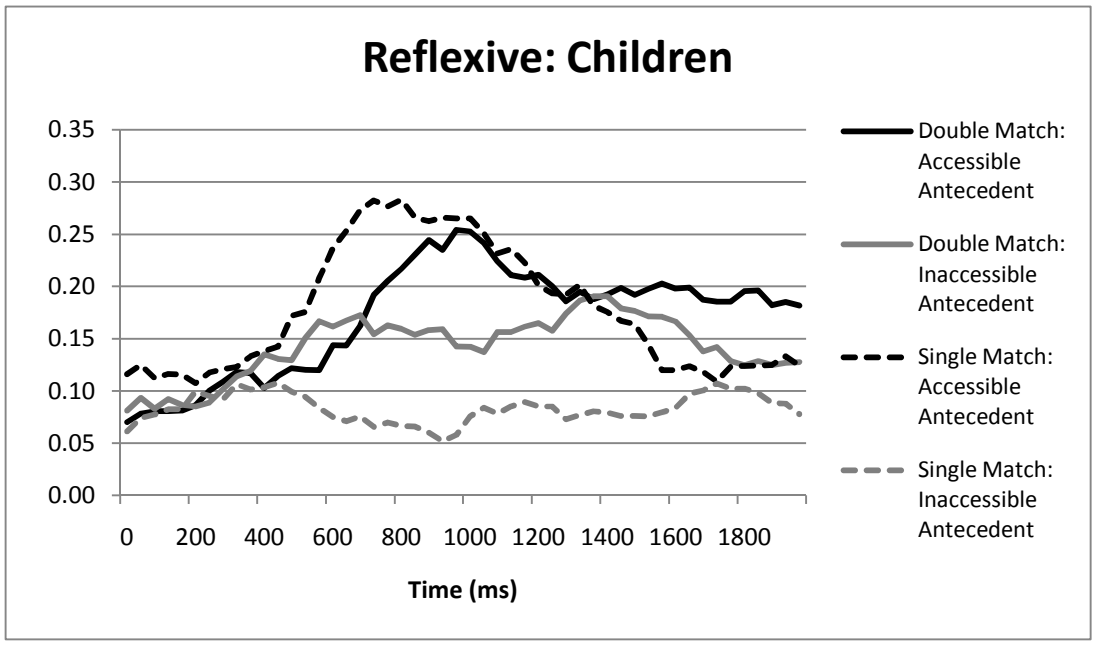


Figure 3: Younger vs. older children's proportions of looks to potential antecedents for reflexives in the double-match condition (subject means).

Figure 3a

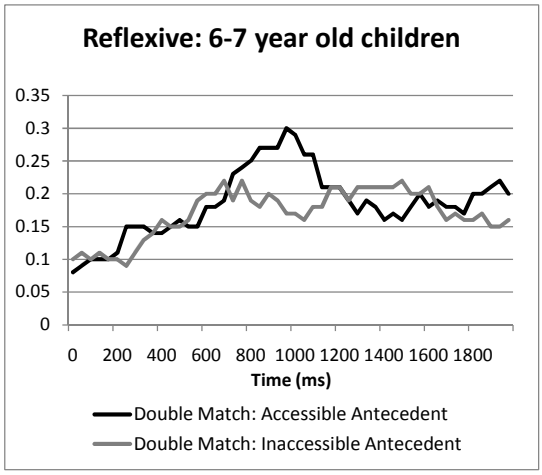


Figure 3b

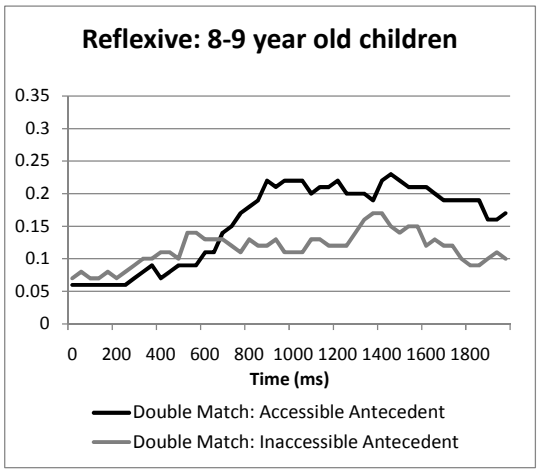


Figure 4: Adults' proportions of looks to potential antecedents on hearing the pronoun (subject means).

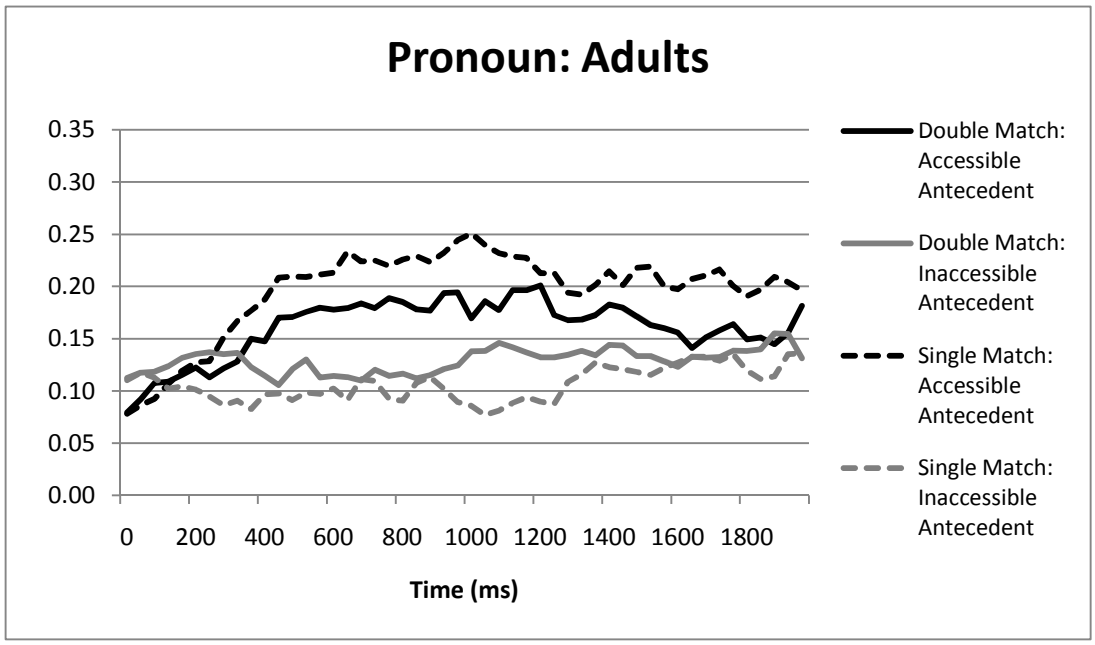


Figure 5: Children's proportions of looks to potential antecedents on hearing the pronoun (subject means).

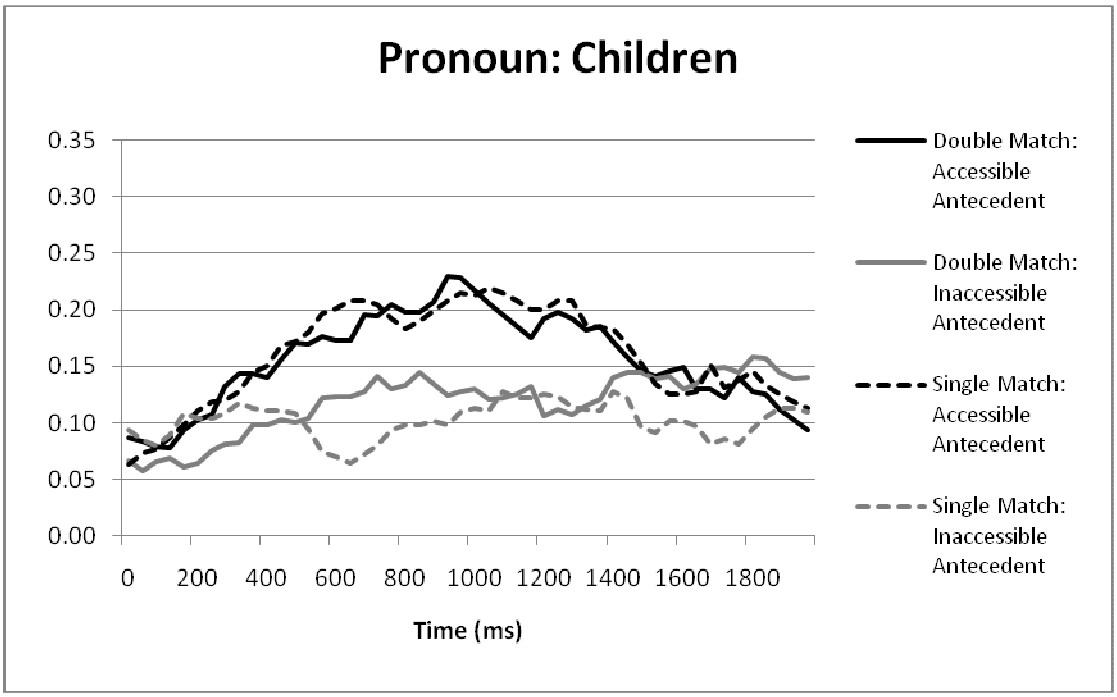


Figure 6: Younger vs. older children's proportions of looks to potential antecedents for pronouns in the double-match condition (subject means).

Figure 6a

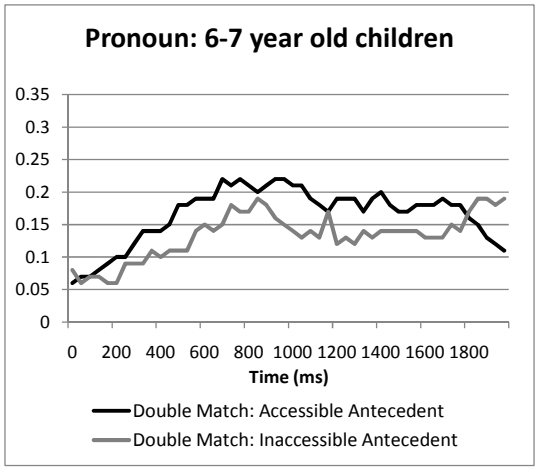


Figure 6b