

# **Adults' on-line comprehension of object pronouns in discourse**

*Petra Hendriks, Arina Banga, Jacolien van Rij, Gisi Cannizzaro & John Hoeks*

## **Introduction<sup>1</sup>**

A well-known finding in the literature on language acquisition is that English-speaking children as old as 6 frequently misinterpret object pronouns as co-referring with the local referential subject. The percentage of errors with respect to this so-called Delay of Principle B Effect (DPBE), however, varies substantially across studies, even when studies are considered that investigate the same language.

Conroy, Takahashi, Lidz and Phillips (2009) showed that in English the DPBE disappears when an elaborate context is presented in which the correct referent and the correct sentence interpretation are made accessible. They conclude from this that English-speaking children possess knowledge of Principle B but are hindered by a discourse context in which the potential referents and interpretations are not appropriately balanced.

A similar disappearance of the DPBE was shown for Dutch by Spenader, Smits and Hendriks (2009). However, rather than presenting children with an elaborate context, they used a short introductory sentence that unambiguously established the correct referent as the discourse topic. They interpret their results as indicating that children's grammar underdetermines the interpretation of pronouns and conclude that Principle B is not a rule of the grammar, but rather is a derived effect resulting from a mature hearer's ability to consider the perspective of the speaker. Because children's interpretations only appear to conform to Principle B if the discourse structure provides a clear topic, Spenader et al. further conclude that children's comprehension of pronouns is sensitive to discourse structure and that children are helped by a coherent discourse. As their analysis is formulated within the framework of Optimality Theory (OT), the relevant discourse conditions are integrated in the grammar.

The different and partly contradictory conclusions of these two studies raise questions regarding English and Dutch children's knowledge of Principle B and the exact contribution of discourse context to pronoun

interpretation. The aim of the present study is to shed more light on the second issue by performing an eyetracking study with Dutch adults. In this study, we test adults' comprehension of object pronouns and reflexives while manipulating the discourse context.

This paper is organized as follows. First, we discuss the off-line studies of Conroy et al. and Spenader et al. with children in more detail. On the basis of these studies, we formulate predictions with respect to the off-line and on-line behavior of adults. We then discuss our eyetracking study with Dutch-speaking adults and present the results of our study. These results are discussed in the light of the formulated hypothesis and predictions. Finally, the implications of our results for the study of anaphora in child language are discussed.

## Theoretical background

### Delay of Principle B Effect

Many experiments in various languages have established that children who correctly interpret reflexives from the age of four or five have trouble interpreting pronouns correctly until the age of 6;6 or even later (e.g., Chien & Wexler, 1990). Consider the following example:

- (1) This is Mama Bear. This is Goldilocks. Is Mama Bear washing herself?
- (2) This is Mama Bear. This is Goldilocks. Is Mama Bear washing her?

Children consistently interpret *herself* in (1) as referring to Mama Bear, thereby showing knowledge of Principle A of Binding Theory, which governs the use and interpretation of reflexives. At the same time, when presented with (2), the same children frequently choose Mama Bear as the referent for *her*. This suggests that they do not yet have knowledge of Principle B of Binding Theory, which governs the use and interpretation of pronouns. This pattern in children's responses is often referred to as the Delay of Principle B Effect (DPBE).

### Discourse matters

Many studies have tried to provide an explanation for the observed asymmetry between children's behavior with Principle A and their behavior with Principle B. One strategy, accepting the observed asymmetry, is to argue that the cause for children's errors with pronouns lies outside the grammar, for example in their lack of pragmatic knowledge (Chien & Wexler, 1990; Grodzinsky & Reinhart, 1993; Thornton & Wexler, 1999) or insufficient processing resources (Reinhart, 2006). Another strategy is to accept this asymmetry but argue that the cause for the asymmetry lies in the properties of the grammar, rather than in extra-grammatical aspects of comprehension (Hendriks & Spenader, 2005/6). A third strategy, adopted by Conroy et al. (2009), is to argue that the observed asymmetry between pronouns and reflexives largely is a reflection of shortcomings of the experimental tests used.

Conroy et al. (2009) carried out three experiments to test the validity of the DPBE as well as the widely assumed asymmetry between quantified antecedents and referential antecedents. The children in their experiments, in which they employed a Truth Value Judgment Task (TVJT), watched an experimenter act out a story with props, and then had to judge whether a statement about the story produced by a puppet, such as "I think that Grumpy painted him", was true or not. The stories were constructed in such a way as to satisfy a number of conditions. First, a potential antecedent for the coreferential interpretation as well as the disjoint interpretation should be available in the discourse (the Availability Assumption). Second, the story should make the correct disjoint interpretation of the pronoun a genuine potential outcome at some point (the Disputability Assumption).

Under these conditions, children made very few errors in the first experiment, accepting the coreferential interpretation in only 11% of trials. Conroy et al. argue that this provides evidence that children know Principle B. In a second experiment, where the pronoun *him* was replaced by the possessive noun phrase *his costume*, children were found to accept reference to the subject in 80% of the trials. According to Conroy et al., this disconfirms the idea that children did so well on the first experiment because they have a general dispreference for bound variable interpretations of pronouns. Reintroducing some of the shortcomings of previous experiments such as Thornton and Wexler's (1999) in the third experiment, children's percentage of incorrect coreferential interpretations increased to 56%. Conroy et al. argue that the results of these three

experiments show that children have knowledge of Principle B but are hindered by a context that is not suitably balanced.

Although children seem to know and generally respect Principle B, they do make more errors in experiments (in roughly 15-30% of trials) than would be expected if Principle B acted as a strong constraint on children's interpretations. This 'residual' DPBE is a real effect, Conroy et al. claim, and may be related to the recent finding in on-line studies of pronoun resolution in adults that adults temporarily consider ungrammatical coreferential antecedents in Principle B contexts (Badecker & Straub, 2002; Kennison, 2003; Runner, Sussman & Tanenhaus, 2003). If these results, obtained by eyetracking and self-paced reading measures, are correct, then Principle B acts as a late filter on the processing of pronouns, after the grammar has generated multiple interpretations (but see Nicol, 1988, who did not find evidence for a coreferential interpretation of the pronoun in her cross-modal priming study; see also Nicol & Swinney, 1989). Because children find it more difficult than adults to inhibit an initial but incorrect interpretation, Conroy et al. argue, this may make them prone to error in their interpretation of pronouns.

Like Conroy et al., Spenader et al. (2009) recognize the importance of the discourse context for a correct assessment of children's knowledge of Principle B. However, rather than implementing the Disputability Assumption in the test materials in the form of an elaborate context story, Spenader et al. implemented this condition as part of their experimental design. In their experiment with Dutch children, the child was told by a puppet that the computer had been built by the experimenter, but the puppet believed that the computer was built wrong. The child was then asked to help repair the computer. So it was made plausible that the pictures and sentences in their Picture Verification Task could - but might not - match. This design allowed Spenader et al. to focus on the other condition that Conroy et al. argue to be crucial in investigating the DPBE, namely the Availability Assumption. In Conroy et al.'s study, the coreferential referent and the disjoint referent were introduced and subsequently referred to in a very elaborate story context, in which these two referents interacted with several other characters. This makes it very difficult to determine the relative salience of the two referents. For this reason, Spenader et al. chose to compare a classic but rather unnatural introduction of the two potential referents, as in Chien and Wexler's (1990) study, with an introduction that is more coherent in terms of Centering Theory (Grosz, Joshi & Weinstein, 1995):

*Classic Condition:*

- (3) Here you see an elephant and an alligator. The elephant is hitting him/himself.

*Single Topic Condition:*

- (4) Here you see an alligator. The elephant is hitting him/himself.

In the Classic Condition, the coreferential and the disjoint referent are introduced in a conjunction and can therefore be taken to be equally salient. As a result, the structure of the linguistic discourse does not provide the listener with any clues as to which of these two referents is to be preferred as the topic of the second sentence, i.e., the test sentence. In the Single Topic Condition, only the disjoint referent is introduced in the first sentence. As a result, only this referent is a potential topic of the test sentence, according to the definitions of Centering Theory. Furthermore, one of the rules of Centering Theory posits that if there is only one pronoun present in the utterance, this pronoun refers to the topic of the utterance (or backward-looking center, in the terminology of Grosz et al., 1995:214). Thus the discourse structure is neutral regarding the interpretation of the pronoun in (3) but promotes the correct interpretation of the pronoun in (4), independently of Principle B.

Overall, in Spenader et al.'s study children's comprehension of reflexives was significantly better than their comprehension of pronouns, consistent with the existence of a DPBE. But whereas children's comprehension of reflexives was similar across conditions, the DPBE was only observed with pronouns in the Classic Condition (31% errors, compared to 14% errors with reflexives in this condition, which was significantly different). In the Single Topic Condition, the DPBE had disappeared completely. Children's comprehension of pronouns in this condition (17% errors) was not significantly different from their comprehension of reflexives (18% errors). Spenader et al. conclude from these results that, for children, pronouns can receive a coreferential as well as a disjoint meaning if the inherent bias of a natural - coherent - discourse context is neutralized. From this, it follows that children are not yet able to apply Principle B. This explanation is compatible with the optimality theoretic (OT) account of the DPBE proposed in Hendriks and Spenader (2005/6). According to this OT account, Principle B is not a constraint of the grammar but rather is a derived or emergent effect. Principle B emerges when hearers optimize bidirectionally and also consider the speaker's perspective. The OT account formalizes the reasoning that if the speaker

would have wanted to express a coreferential meaning, the best option for the speaker according to the constraints of the grammar would have been to use a reflexive. If the speaker did not use a reflexive but rather used a pronoun, the hearer may conclude that the coreferential meaning is not the meaning intended by the speaker, and hence block this interpretation for the pronoun. Assuming that children are not yet able to take into account the speaker's perspective in their linguistic optimization (cf. de Hoop & Krämer, 2005/6; Hendriks & Spenader, 2005/6), this explains why children do not discard the coreferential interpretation for a pronoun. At the same time, the OT grammar predicts that children's production of pronouns in the Classic Condition is adult-like. This latter prediction was confirmed by the results of Spenader et al.'s elicited production task.<sup>2</sup>

So, why do children perform so much better with pronouns in the Single Topic Condition than in the Classic Condition? In the Single Topic Condition, as in all other contexts where the disjoint referent is much more salient than the coreferential referent, a general preference for pronouns to refer to the discourse topic may facilitate selection of the disjoint referent over the coreferential referent. Spenader et al. suggest that this preference is incorporated in the grammar as a violable constraint, which we refer to here as ProTop. Assuming that the more highly ranked constraints of the grammar do not decide between a coreferential and a disjoint meaning for a pronoun, this weaker constraint ProTop will determine the selection of the antecedent. In optimality theoretic terms, this is known as 'the emergence of the unmarked': usually only the effects of stronger constraints are visible within a language, but in particular circumstances a weaker constraint becomes crucial. Because ProTop promotes selection of the discourse topic as the antecedent of a pronoun, its application results in selection of the correct antecedent in the Single Topic Condition in Spenader et al.'s study. This is because, in their study, the discourse topic introduced in the first sentence always is the correct antecedent. In the Classic Condition, on the other hand, the constraint ProTop will lead to guessing behavior. Because the coreferential and the disjoint referent are equally salient in the first sentence, ProTop will not be able to decide between these two referents. Hence, children will simply select one of these referents at chance.

In contrast to children, adults are not dependent on the presence of contextual cues to arrive at the correct interpretation of a pronoun. Rather, they are able to discard the coreferential meaning by reasoning about the speaker's alternative linguistic options. Therefore, they select the disjoint referent also in the Classic Condition.

### Adults' processing of pronouns

According to Conroy et al. (2009), there are two different sources for children's DPBE errors: (1) an unbalanced discourse context, and (2) a failure to inhibit the initially activated but incorrect coreferential interpretation. Only children are hindered by an unbalanced context. Adults have no problems applying Principle B in these situations. Conroy et al. do not provide an explanation for why adults cannot be contextually coerced into an ungrammatical interpretation. However, an obvious explanation, which is compatible with their account of the DPBE, is that adults' application of Principle B is independent of discourse context. Spender et al. (2009), on the other hand, argue that Principle B is a derived effect which requires that hearers take into account the speaker's perspective. Although children are unable to derive Principle B, a preference for the pronoun to refer to the discourse topic may nevertheless direct children toward the correct interpretation. This preference is argued to have the form of a weak constraint that is part of the grammar. But if this constraint is part of the grammar, its effects may show up in adults' on-line processing as well.

On the basis of these two different positions regarding the DPBE, we can formulate a number of predictions regarding adults' and children's on-line processing of pronouns. Both for adults and for children, we predict that pronouns are initially ambiguous and activate the coreferential as well as the disjoint referent. This is in line with results of earlier processing studies (Badecker & Straub, 2002; Kennison, 2003; Runner, Sussman & Tanenhaus, 2003) as well as with the theoretical assumptions of the two accounts under consideration. If a preference for pronouns to refer to the topic is a constraint of the grammar, as Spender et al. contend, we expect the effects of this preference not to be confined to children's off-line responses. Rather, this preference may also show up in children's and adult's on-line processing. In particular, we predict that children as well as adults will experience more processing difficulty interpreting pronouns in the Classic Condition than in the Single Topic Condition. Because this preference only pertains to the interpretation of pronouns, we expect no effects of context when they process reflexives. In contrast, a prediction that seems compatible with Conroy et al.'s account of the DPBE is that effects of context are observable neither in adults' on-line processing of reflexives nor in their on-line processing of pronouns. In this study, we focus on adults' on-line processing and investigate the predictions

regarding adults' on-line comprehension of pronouns and reflexives. In the next section, we describe the details of our eyetracking experiment with Dutch adults, in which we use similar materials to Spender et al. (2009).

## Experiment

### Participants

Twenty-five adult native speakers of Dutch participated in this study. Most of them were university students, who participated voluntarily. One participant was excluded from data analysis because his other native language was Frisian, and pronouns in Dutch and Frisian have different properties. Therefore, our analyses are based on the data of the resulting 24 participants: 13 men and 11 women (mean age 22 years, age range 18-27).

### Materials and design

Each item consisted of a pre-recorded sequence of two sentences in combination with a picture. The sequences of sentences consisted of an introductory sentence followed by the test sentence. The pictures displayed two animals of approximately equal size engaged in a self-oriented or other-oriented action (see Figure 1).



*Figure 1.* A picture displaying a self-oriented action (left) and a picture displaying an other-oriented action (right), adapted to black and white print.

The recorded sentences were spoken by a female student, who had a neutral Dutch accent. The first sentence of each sentence pair served as an introduction of the character(s) and the setting, whereas the second sentence contained an anaphor which referred back to one of the two referents introduced in the previous sentence, or to the single referent introduced in the previous sentence. The following verbs were used in the

second sentence, which was the test sentence: *aankleden* 'to dress', *bijten* 'to bite', *kietelen* 'to tickle', *schilderen* 'to paint' / *tekenen* 'to draw', *slaan* 'to hit', *vastbinden* 'to tie', *wijzen naar* 'to point at' and *schminken* 'to make up'.

Two factors were manipulated in the experimental items. The first factor was Type of Introductory Sentence: The introductory sentence introduced both animals (Classic Condition, or *C*) or only one of the animals (Single Topic Condition, or *S*). The second factor was Type of Anaphor: The sentences contained a pronoun (*P*) or reflexive (*R*). An example of each of the four resulting experimental conditions is given in (5)-(8).

*Classic Condition+Pronoun (CP):*

- (5) Een aap en een schildpad zijn op het strand. De aap kietelt hem.  
'A monkey and a turtle are on the beach. The monkey is tickling him.'

*Classic Condition+Reflexive (CR):*

- (6) Een aap en een schildpad zijn op het strand. De aap kietelt zichzelf.  
'A monkey and a turtle are on the beach. The monkey is tickling himself.'

*Single Topic Condition+Pronoun (SP):*

- (7) Een schildpad is op het strand. De aap kietelt hem.  
'A turtle is on the beach. The monkey is tickling him.'

*Single Topic Condition+Reflexive (SR):*

- (8) Een schildpad is op het strand. De aap kietelt zichzelf.  
'A turtle is on the beach. The monkey is tickling himself.'

Four versions of the experiment were constructed by a Latin square design, so that each list contained four items of each condition and one version of each item. This was done by using each of the 8 verbs twice, but with different pairs of animals. As a result, each version of the experiment consisted of 16 experimental items, distributed equally over four blocks. In addition, to distract the attention of the participants from the goal of the experiment, 32 filler items were included which were also preceded by an introductory sentence but did not contain an anaphor. Of these fillers, 18 contained a transitive verb and a definite description as the direct object, and 14 contained an intransitive verb. See Banga (2008) for all four lists as well as a detailed description of the distribution of experimental items and fillers over the lists. For half of the experimental items, the test sentence and the picture matched, whereas for the other half of the items the test

sentence and the picture did not match with respect to the orientation of the action (self-oriented vs. other-oriented). Also for the fillers, half of the sentences and pictures matched, whereas the other half did not match with respect to the meaning of the predicate. The matching items are expected to lead to a ‘yes’ response by participants, whereas the mismatching items are expected to lead to a ‘no’ response.

The similarities and differences between items in the four conditions are listed in Table 1. The pictures for pronouns in a match situation and reflexives in a mismatch situation display other-oriented actions. The pictures for pronouns in a mismatch situation and reflexives in a match situation display self-oriented actions. The sentences with pronouns and reflexives in the same context condition are identical until the onset of the anaphor.

Table 1. Specification of the experimental conditions.

Condition	Action in picture	Context sentence	Anaphor
CP-Match	other-oriented	Classic	pronoun
CP-Mismatch	self-oriented	Classic	pronoun
SP-Match	other-oriented	Single Topic	pronoun
SP-Mismatch	self-oriented	Single Topic	pronoun
CR-Match	self-oriented	Classic	reflexive
CR-Mismatch	other-oriented	Classic	reflexive
SR-Match	self-oriented	Single Topic	reflexive
SR-Mismatch	other-oriented	Single Topic	reflexive

To obtain eye-movement data for the correct and the incorrect referent, two areas of interest (AoI) were defined in each picture (see Figure 2).



Figure 2. Defining the areas of interest (AoIs). AoIs were drawn by hand at approximately 1 cm distance around the animals on the pictures.

The first AoI is the referent that is the subject of the test sentence (NP1), which also is the patient of a self-oriented action. The second AoI is the other referent (NP2), which is the patient of an other-oriented action or the bystander in case of a self-oriented action. The borders of the AoIs were drawn by hand at approximately 1 cm distance around the two animals on the pictures, except when the animals (almost) touched each other. In that case the borders of the AoIs were drawn closer to the edges of the animals, so that there would be no overlap between the AoIs.

### Procedure

Using a remote Tobii T120 eyetracker in combination with two computers, we measured adults' accuracy, reaction times and eye movements during a Picture Verification Task. One computer with E-Prime (Schneider, Eschman & Zuccolotto, 2002) was used for stimuli presentation and collection of the accuracy and reaction time data. The other computer with Tobii Studio software was used for collecting the gaze data (in this study, at a frame rate of 60 Hz). The remote Tobii eyetracker is integrated in a 17 inch TFT monitor, has no visible or moving tracking devices, and allows a freedom of head movement of 44 x 22 x 30 cm. The eye-movement data reported are an average of both eyes.

Participants were tested individually in a room without any windows, in which the eyetracker and associated equipment were located. They were seated in front of the Tobii eyetracker monitor, with a keyboard attached to the E-Prime computer placed before them. The experimental procedure involved four parts. The first part was the calibration, the second part was a training session of three items through which the participants were familiarized with the experimental procedure, the third part was the actual experiment, and the fourth part was a questionnaire requesting some basic personal information that the participant was asked to complete. Men and women were equally distributed over the four different versions of the experiment.

The participants were told that they were going to listen to a series of short stories while looking at pictures on the computer screen. The picture always appeared on the screen first, followed by the start of the sound file after approximately 1000 ms. The participants were asked to register as fast and as accurately as possible whether the second sentence of the story they heard corresponded to the picture they saw. If it did, the participants had to

press the ‘Q’ key on the keyboard; if it did not, they had to press the ‘P’ key. The entire test session lasted about 25 minutes.

## Results

### Responses

Although it is to be expected that adults hardly make any errors in the Picture Verification Task, the accuracy of responses was analyzed to confirm this expectation.<sup>3</sup> In total 335 responses were given, of which 322 (96%) were correct and only 13 (4%) were incorrect. For the four conditions, the proportions of correct responses out of the total amount of responses were calculated for each participant and for each item. The mean proportions of correct responses and their standard deviations, all based on participant analysis, are listed in Table 2.

Table 2. Mean proportion of correct responses (standard deviations), in the Picture Verification Task.

Pronoun		Reflexive	
Classic	Single Topic	Classic	Single Topic
0.93 (0.14)	0.99 (0.07)	1.00 (0.00)	0.91 (0.18)

The proportions were first arcsine-transformed. Repeated Measures ANOVAs were then run on these transformed proportions with Type of Introductory Sentence (classic versus single topic) and Type of Anaphor (pronoun versus reflexive) as within-participants and within-items factors. No significant main effects were found for Type of Introductory Sentence, or Type of Anaphor (all  $F$ -values  $< 1$ ). There was also no significant interaction of Type of Introductory Sentence x Type of Anaphor ( $p$ -values  $> .10$ ). So the few errors that were made were equally distributed across conditions.<sup>4</sup>

### Reaction times

In addition to their responses in the Picture Verification Task, we also measured participants’ reaction times on the task. Differences in reaction times between conditions are generally considered to be an indication of

differences in processing difficulty. If a preference for pronouns to refer to the topic is part of the grammar (cf. Spenser et al., 2009), we expect pronouns to take more time to be interpreted in the Classic Condition than in the Single Topic Condition. If, on the other hand, context does not play a role in adults' on-line processing of pronouns (a possibility compatible with Conroy et al., 2009), we do not expect any significant differences between the two conditions. For reflexives, both accounts predict that there will not be any differences between the two conditions.

Only correct responses were included in the analysis of reaction times. Although no accuracy differences were found between the four conditions, this was done to rule out the possibility that incorrect responses may have influenced reaction times. For the four conditions, the mean reaction times and the standard deviations were calculated for each participant and for each item. Reaction times were measured from the onset of the anaphor (i.e., the disambiguating word *hem* 'him' or *zichzelf* 'himself') until the response given. The mean reaction times and standard deviations are listed in Table 3 and are graphically presented in Figure 3. Numbers are based on participant analysis.

Table 3. Mean reaction times in milliseconds (standard deviations), measured from onset of anaphor.

Pronoun		Reflexive	
Classic	Single Topic	Classic	Single Topic
1642 (473)	1301 (364)	1319 (355)	1290 (327)

Repeated Measures ANOVAs were run on the reaction times with Type of Introductory Sentence (classic versus single topic) and Type of Anaphor (pronoun versus reflexive) as within-participants and within-items factors. The main effect of Type of Introductory Sentence was significant by participants, but did not reach significance in the analysis by items,  $F_1(1,23) = 8.57$ ,  $MSE = 78391.010$ ,  $p = .008$ ;  $F_2(1,13) = 1.71$ ,  $MSE = 92937.804$ ,  $p = .213$ . Similarly, the effect of Type of Anaphor was significant by participants but not by items,  $F_1(1,23) = 8.134$ ,  $MSE = 100829.805$ ,  $p = .009$ ;  $F_2(1,13) = 2.93$ ,  $MSE = 88210.522$ ,  $p = .111$ . These effects, however, were qualified by a significant interaction of Type of Introductory Sentence x Type of Anaphor,  $F_1(1,23) = 6.89$ ,  $MSE = 84236.47$ ,  $p = .015$ ;  $F_2(1,13) = 11.500$ ,  $MSE = 0.010$ ,  $p = .005$ . Paired  $t$ -tests with Bonferroni correction showed that responses to pronouns in the

14 Petra Hendriks, Arina Banga, Jacolien van Rij, Gisi Cannizzaro & John Hoeks

Classic Condition took longer than responses in any of the other three conditions (all  $p$ -values < .005). Reaction times in these latter three conditions were not significantly different.

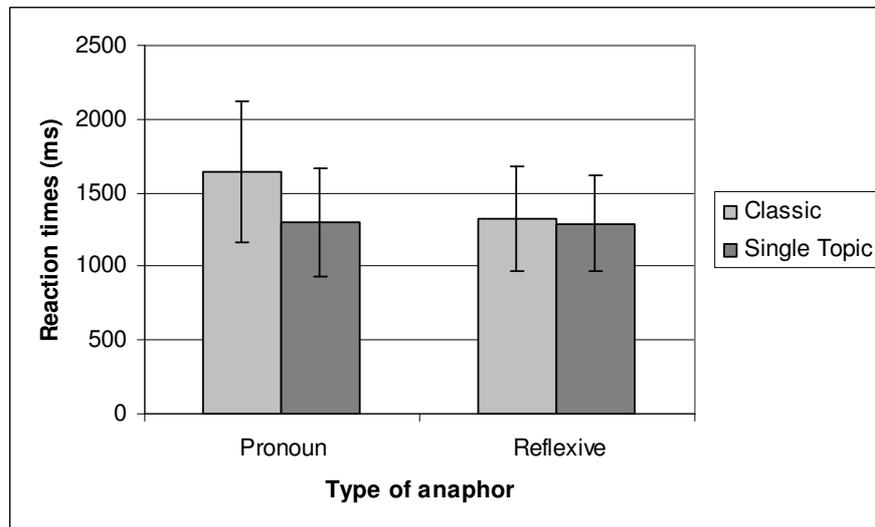


Figure 3. Mean reaction times measured from onset of anaphor.

#### Eye movements

In addition to collecting reaction times, we recorded the eye movements of the participants, as the timing and pattern of looks to potential referents has been argued to provide information about which referents are considered during the comprehension of pronouns and reflexives (Runner, Sussman & Tanenhaus, 2003; Sekerina, Stromswold & Hestvik, 2004).

Figures 4-7 below present the graphical results for the four conditions, distinguishing between match situations and mismatch situations. The figures show the proportion of fixations averaged over participants from the presentation of the picture until after a response has been given following the second sentence. The average duration per condition of the introductory sentence is represented by the left horizontal bar above the graph, and the average duration per condition of the second sentence (the test sentence) by the right horizontal bar. The onset of the anaphor is indicated by 'A' and

the mean reaction time by 'RT'. The target is always the correct referent, and the distractor the other referent. The category 'other' includes all looks outside these two Areas of Interest.

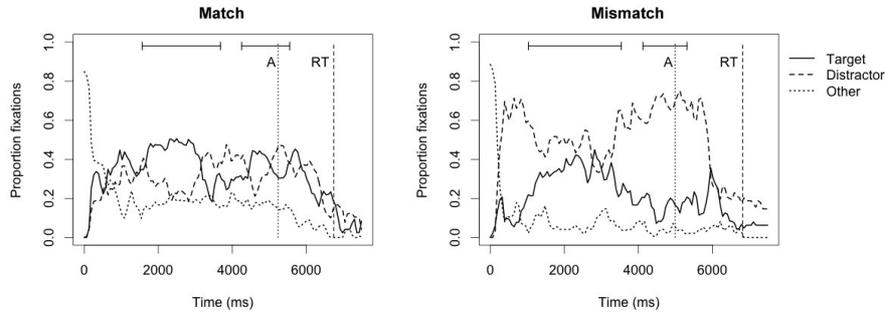


Figure 4. Pronouns in Classic Condition (CP).

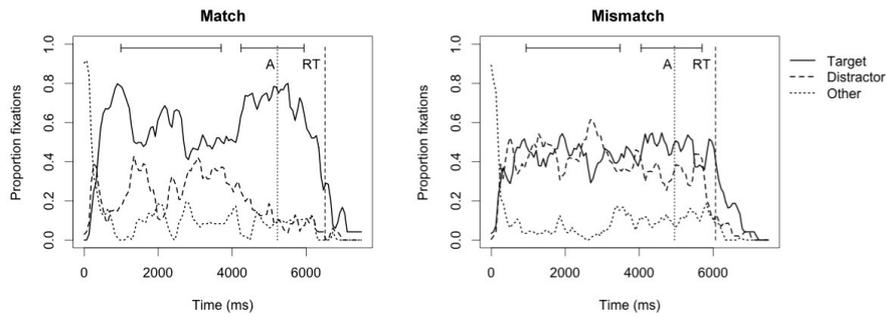


Figure 5. Reflexives in Classic Condition (CR).

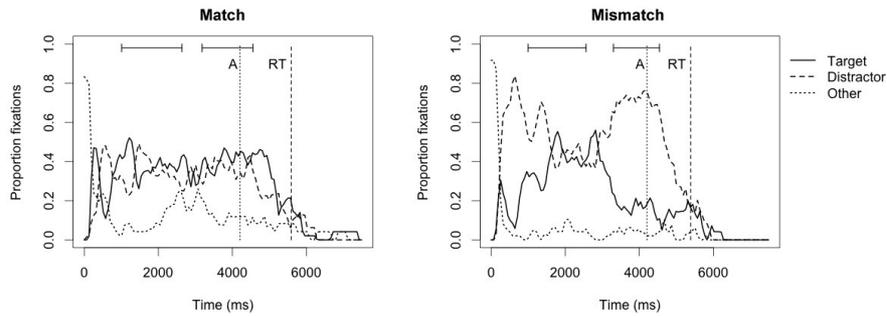


Figure 6. Pronouns in Single Topic Condition (SP).

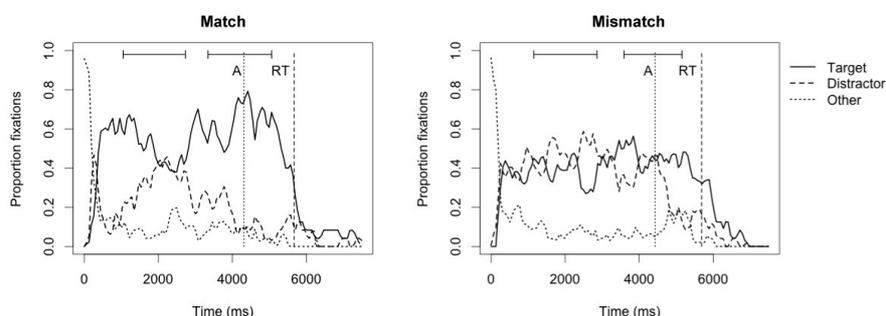


Figure 7. Reflexives in Single Topic Condition (SR).

As can be seen from Figures 4-7, the pictures have a considerable effect on the eye movements. If the picture presents an other-oriented action, as with pronouns in the match conditions and reflexives in the mismatch conditions, the pattern of looks is very similar. In these cases, participants looked at the two referents approximately equally often, somewhat more than 40% of times. This pattern is clearly distinct from the pattern that can be observed with pictures presenting a self-oriented action. For pronouns in the mismatch conditions as well as reflexives in the match conditions, more looks (almost 60%) are to the agent of the self-oriented action, which is at the same time the patient of the action. Fewer looks (roughly 30%) are to the other referent. This effect of the pictures results in a dominance of looks to the correct referent (the target) for the reflexive but to the incorrect referent (the distractor) for the pronoun.

In addition to these general looking patterns, we considered two specific measures in the participant's eye-movement data: (1) mean proportions of observation length to the correct referent, and (2) mean time to first fixation on the correct referent. Observation length is an overall measure of sentence interpretation that aggregates all looking times to a given referent (or to be more precise: within the area of interest defined for that referent) from the onset of the anaphor until the participant has given a response. The onset of the anaphor is the disambiguating point in the sentence and is in most items the last word of the sentence.<sup>5</sup> Because of individual differences in reaction times (i.e., time between onset of anaphor and actual response), observation lengths are normalized by dividing the observation length for the correct referent by the sum of the observation length for the correct referent and the observation length for the incorrect referent. If participants look less at the correct referent in a particular condition as

compared to another condition, we may assume that they find the correct referent a less probable antecedent for the anaphor in this condition than in the other condition. The second measure, mean time to first fixation on the correct referent, is the time in milliseconds from the onset of the anaphor until the start of the first fixation on the correct referent (or to be more precise: within the area of interest defined for the correct referent). It is an early measure of interpretation that yields an indication of how much time it took the participant to zoom in on the correct referent for the first time after the onset of the anaphor. The faster this process, the easier accessing the correct interpretation seems to be. So if participants take less time to fixate on the correct referent in a particular condition as compared to another condition, we may assume that they find it easier to access the correct referent in this condition.

Mean proportions of observation length are listed in Table 4 and are graphically presented in Figure 8.

Table 4. Mean proportions of observation length (standard deviations), measured from onset of anaphor until correct response.

Pronoun		Reflexive	
Classic	Single Topic	Classic	Single Topic
0.39 (0.16)	0.43 (0.34)	0.73 (0.18)	0.69 (0.12)

The proportions were first arcsine-transformed. Repeated Measures ANOVAs were then run on these transformed proportions with Type of Introductory Sentence (classic versus single-topic) and Type of Anaphor (pronoun versus reflexive) as within-participants and within-items factors. There was no main effect of Type of Introductory Sentence (both  $F$ -values  $< 1$ ), but a main effect of Type of Anaphor was found,  $F_1(1,21) = 41.51$ ,  $MSE = 0.273$ ,  $p < .001$ ;  $F_2(1,15) = 17.01$ ,  $MSE = 0.553$ ,  $p < .001$ . In the reflexive conditions ( $M = 0.71$ ), higher proportions of observation length were to the intended referent compared to the pronoun conditions ( $M = 0.41$ ). No interaction between Type of Introductory Sentence x Type of Anaphor was found ( $p$ -values  $> .30$ ).

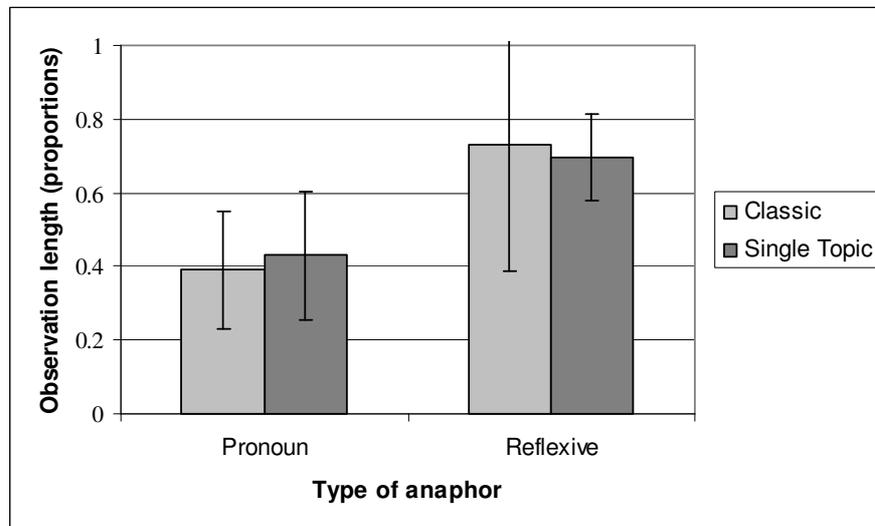


Figure 8. Mean proportions of observation length, measured from onset of anaphor until correct response.

The mean times to first fixation on the correct referent were also analyzed. These data are listed in Table 5 and are graphically presented in Figure 9.

Table 5. Mean time to first fixation in milliseconds (standard deviations), measured from onset of anaphor.

Pronoun		Reflexive	
Classic	Single Topic	Classic	Single Topic
358 (278)	323 (278)	73 (87)	85 (103)

Repeated Measures ANOVAs were run on the mean times to first fixation on the correct referent with Type of Introductory Sentence (classic versus single-topic) and Type of Anaphor (pronoun versus reflexive) as within-participants and within-items factors. No main effect of Type of Introductory Sentence emerged (both  $F$ -values  $< 1$ ), but a main effect of Type of Anaphor was found,  $F_1(1,22) = 39.852$ ,  $MSE = 39387.44$ ,  $p < .001$ ;  $F_2(1,12) = 15.77$ ,  $MSE = 65886.81$ ,  $p = .002$ . For the reflexive conditions ( $M = 79$  ms), the time to the first fixation on the correct referent was shorter than for the pronoun conditions ( $M = 340$  ms). No interaction between Type of Introductory Sentence x Type of Anaphor was found ( $F$ -values  $< 1$ ).

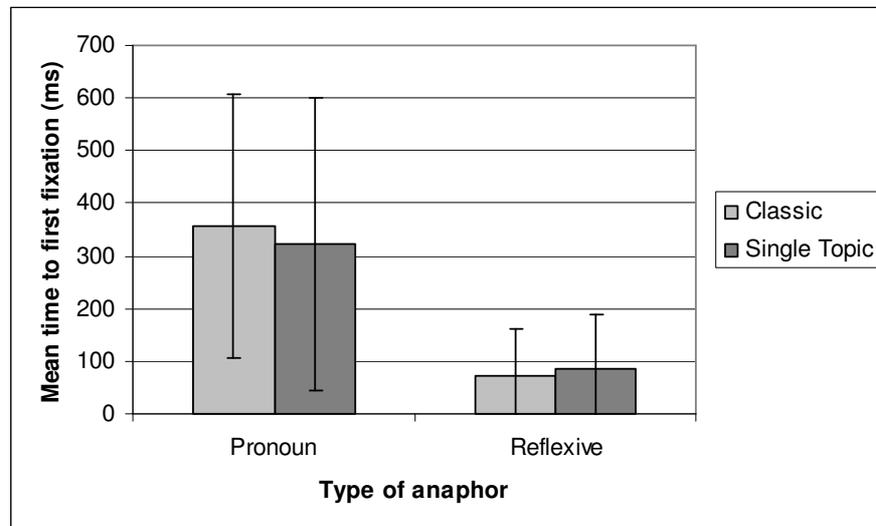


Figure 9. Mean time to first fixation, measured from onset of anaphor.

## Discussion

### Adults' on-line comprehension of anaphora

The central question of our study is whether and how discourse context influences adults' on-line comprehension of pronouns. On the basis of the literature, two competing hypotheses were formulated. The first hypothesis, compatible with Conroy et al.'s (2009) partly reductionist account of the DPBE, is that discourse context has no significant effects on adults' on-line comprehension of pronouns. The second hypothesis, derived from Spenader et al.'s (2009) optimality theoretic explanation of the DPBE, is that constraints on local discourse coherence not only have effects on children's off-line responses but also on adults' on-line processing of pronouns.

We looked at adults' accuracy, reaction times and eye movements during a Picture Verification Task. As both accounts would predict, the adult participants in our study hardly made any comprehension errors. However, their reaction times with pronouns in the Classic Condition were significantly longer than their responses in the other three conditions. As we found a significant interaction, and not just main effects of either

linguistic information or visual information, it is rather unlikely that these reaction times are mere reflections of processes having to do with decision making. Instead, we assume that these differences in reaction times to a large extent reflect differences in the complexity of processes underlying sentence comprehension. This would indicate that the adults experienced more processing difficulties with pronouns in the Classic Condition than in the Single Topic Condition. Thus, the results of our on-line study with adults appear to be consistent with the results of Spenader et al.'s off-line study with children. They found that children make more errors on pronoun interpretation in the Classic Condition than in the Single Topic Condition. Because children's errors and adults' processing difficulties occur in the same type of sentence, it is not implausible that they stem from the same source. That is, adults' longer reaction times for pronouns in the Classic Condition may provide support for Spenader et al.'s suggestion that the observed preference for pronouns to refer to the topic is part of the grammar.

Participants' eye movements however showed a different picture. The two specific measures we looked at only showed a main effect of type of anaphor: Participants looked for a shorter time at the correct referent for the pronoun than for the reflexive, and it took them longer to fixate on the correct referent for the pronoun than for the reflexive. Surprisingly, the two measures showed no effects of context. Participants did not look for a shorter time at the correct referent for the pronoun in the Classic Condition than in the Single Topic Condition, nor did it take them longer to fixate on the correct referent for the first time. These results then seem to support the account of Conroy et al., since adults do not appear to be influenced by discourse context in their looking behavior when interpreting pronouns or reflexives. So the reaction times appear to support an analysis according to which discourse context has a significant effect on adults' processing of object pronouns, whereas the eye-movement data appear to support an analysis according to which discourse context has no effects at all. How are we to reconcile these different results?

Generally speaking, eye movements do not seem to be as strongly linked to processing difficulty as reaction times. It is possible that the eye movements in our study have been influenced by the task. This possibility is supported by the participants' patterns of looking to the correct and incorrect referent over time. As can be seen from Figures 4-7, the pictures had a considerable effect on the eye movements. The effects of the pictures are large enough perhaps to have masked potential context effects in the

eye-movement data. On all numerical measures, pronouns in the Classic Condition seem most difficult for participants. Pronouns in the Classic Condition gave rise to the longest reaction times, the shortest observation lengths to the correct referent, and the longest times to first fixation on the correct referent. Although only the result from reaction times is significant, the tendencies displayed in the eye-movement data are certainly not incompatible with the result from reaction times.

If participants' processing difficulties in our study are more accurately reflected by their reaction times than by their eye movements, this has implications for the use of eyetracking in the study of language. Researchers have used eyetracking in combination with a wide array of tasks, such as Picture Selection Tasks (Sekerina et al., 2004), action based tasks (Runner et al., 2003), and Truth Value Judgment Tasks and Picture Verification Tasks (Arnold et al., 2000). The general consensus seems to be that eyetracking is less sensitive to task effects than other on-line or off-line methods, because of the underlying assumption that "the mind is going where the eye is going" (Trueswell & Gleitman, 2007). However, our study suggests that Picture Verification Tasks, but perhaps also other psycholinguistic tasks, may influence participants' eye movements to a considerable extent.

#### Implications for child language research

The present study investigated adults' on-line processing of pronouns and reflexives. It revealed certain similarities between adults' on-line processing and children's off-line interpretation. These similarities indicate that for children as well as adults pronouns in the Classic Condition are more difficult to interpret than pronouns in the Single Topic Condition or reflexives in either condition. These results follow from the predictions of Spenader et al. (2009), who assume that children are not yet able to apply Principle B of Binding Theory but use cues from discourse context whenever they can to interpret pronouns. Adults' reaction times for pronouns in the Single Topic Condition were comparable to their reaction times for reflexives in either condition. This suggests that adults are helped by a coherent discourse too.

Our study also displayed differences between adults' on-line processing and children's off-line interpretation. Whereas children's off-line interpretation of object pronouns is heavily influenced by discourse

context, as was shown by Conroy et al. (2009) and Spenader et al. (2009), we did not find any effects of discourse context in the eye-movement data of the adults. The results of the adults may follow from the account of Conroy et al. On the basis of the off-line results of their experiments, Conroy et al. conclude that children are hindered by an unbalanced context. The adult controls in their experiments, on the other hand, did not seem to be hindered by an unbalanced context in their off-line responses. If the effects of context are assumed to be caused by factors related to the Truth-Value Judgment Task that disappear with age, it is predicted that adults are not influenced by the discourse context in their on-line processing either. However, we suggested that the absence of context effects in adults' eye-movement data may have been caused by particular task effects associated with the Picture Verification Task.

To make more sense of the different conclusions that can be drawn from adults' reaction times and their eye movements, and to shed more light on children's knowledge of Principle B, it would be useful to study children's on-line processing of pronouns and reflexives in relation to the structure of the discourse. If children show the same pattern of eye movements as the adults in our study, this could suggest that eye movements are not directly tied to a participant's interpretation of an anaphor and may reflect a different set of (task- or materials-induced) processes. On the other hand, differences between children's eye movements and adults' eye movements may point at a different contribution of discourse context in children's and adults' interpretation of pronouns.

## **Conclusions**

In this study we tested Dutch adults on a Picture Verification Task while recording their reaction times and monitoring their eye movements. As we were interested in the effects of discourse context on the interpretation of object pronouns and reflexives, we manipulated the structure of the introductory sentence. Although the adults in our study hardly made any comprehension errors, their reaction times were significantly slower when the introductory sentence did not unambiguously establish a discourse topic. This suggests that the structure of the discourse context is important for pronoun interpretation and influences both children's off-line interpretation of object pronouns (resulting in DPBE errors in particular contexts) and adults' on-line processing of object pronouns (resulting in

slower reaction times in the same contexts). These results seem to be at odds with approaches that try to reduce the DPBE to an experimental artifact. On the other hand, adults' eye movements did not provide significant evidence for effects of discourse context on their processing of object pronouns. We believe that this may have been caused by task effects.

## Notes

1. Contact information Petra Hendriks, Jacolien van Rij, Gisi Cannizzaro and John Hoeks: CLCG, University of Groningen, P.O. Box 716, 9700 AS Groningen, The Netherlands, P.Hendriks@rug.nl, J.C.van.Rij@rug.nl, C.L.Cannizzaro@rug.nl, J.C.J.Hoeks@rug.nl. Contact information Arina Banga: CLS, Radboud University Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands, arina.banga@mpi.nl. This investigation was supported by a grant from the Netherlands Organisation for Scientific Research, NWO, awarded to Petra Hendriks (grant no. 277-70-005) for the VICI project "Asymmetries in Grammar". The authors thank Robbert Prins and Petra van Berkum for drawing the pictures used in the experiment, and Colin Phillips, Jennifer Spenader, the editors Cornelia Hamann and Esther Ruigendijk, and three anonymous reviewers for their useful comments and discussion.
2. The reviewers suggest that a potential problem for the analyses of both Conroy et al. and Spenader et al. may be the observed cross-linguistic differences in the DPBE. Although the DPBE has been observed in simple transitive sentences in English and Dutch, this effect does not seem to arise in similar sentences in the closely related language German (see Ruigendijk, 2008). These cross-linguistic differences are problematic for Conroy et al.'s account, which appeals to processing factors for the 'residual' DPBE effect. It is highly unlikely that speakers of German are better at inhibiting incorrect interpretations than speakers of Dutch or English. However, cross-linguistic differences do not seem to be a priori incompatible with Spenader et al.'s OT account of the DPBE. In OT, languages are assumed to differ as a result of a different ranking of the constraints of the grammar, which may give rise to different lexical inventories and different locality conditions on reflexive binding. Because acquisition delays are predicted to arise in those cases where unidirectional optimization yields a different result than bidirectional optimization when using the same grammar, different grammars may yield different predictions regarding the DPBE.
3. Because accidentally, for one item in all conditions, the wrong audio file and picture were combined (resulting in a mismatch item rather than a match item, but with an incorrect agent rather than an incorrect patient), this item (Item 2)

24 Petra Hendriks, Arina Banga, Jacolien van Rij, Gisi Cannizzaro & John Hoeks

was removed from our analysis. Another item (Item 15) was removed from our analysis of the SP condition only because the verb/action for 'to dress' was used rather than for 'to make up'.

4. Although the use of a definite article without previous mention of the referent is generally believed to be infelicitous, the fact that participants readily accepted sentences such as (7) and (8) indicates that visual information can render a referent sufficiently familiar to license the use of a definite article.
5. Two of the verbs that were used require a particle (*aankleden* 'to dress' and *vastbinden* 'to tie'). These particles (*aan* and *vast*) are usually placed in sentence-final position in Dutch main clauses and hence follow the anaphor in the test sentences. However, as these verbs were distributed equally across conditions, their inclusion did not influence the results.

## References

- Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C.  
2000 The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition* 76, B13-B26.
- 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.
- Banga, A.  
2008 *On-line measures in adults' comprehension of anaphors in coherent and incoherent discourse*. MA thesis, University of Groningen.
- Chien, Y. C. & Wexler, K.  
1990 Children's knowledge of locality conditions on binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* 1, 225-295.
- Conroy, S., Takahashi, E., Lidz, J. & Phillips, C.  
2009 Equal treatment for all antecedents: How children succeed with Principle B. *Linguistic Inquiry* 40, 446-486.
- de Hoop, H. & Krämer, I.  
2005/6 Children's optimal interpretations of indefinite subjects and objects. *Language Acquisition* 13, 103-123.
- Grodzinsky, Y. & Reinhart, T.  
1993 The innateness of binding and the development of coreference. *Linguistic Inquiry* 24, 69-101.
- Grosz, B., Joshi, A., & Weinstein, S.  
1995 Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics* 21, 203-225.

- Hendriks, P. & Spenader, J.  
2005/6 When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition* 13, 319-348.
- Kennison, S.  
2003 Comprehending the pronouns “her”, “him” and “his”: Implications for theories of referential processing. *Journal of Memory and Language* 49, 335-352.
- Nicol, J. L.  
1988 *Coreference processing during sentence comprehension*. Ph.D. Dissertation, MIT.
- Nicol, J. L. & Swinney, D.  
1989 The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18, 5-19.
- Reinhart, T.  
2006 *Interface Strategies: Optimal and Costly Computations*. Cambridge, MA: MIT Press.
- Ruigendijk, E.  
2008 Pronoun interpretation in German kindergarten children. In: A. Gavarró & M. João Freitas (Eds), *Proceedings of GALA 2007*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Runner, J. T., Sussman, R. S. & Tanenhaus, M. K.  
2003 The influence of Binding Theory on the on-line reference resolution of pronouns. *Proceedings of NELS*.
- Schneider, W., Eschman, A., & Zuccolotto, A.  
2002 *E-Prime User's Guide*. University of Pittsburgh Psychology Software Tools, Inc.
- Sekerina, I. A., Stromswold, K. & Hestvik, A.  
2004 How do adults and children process referentially ambiguous pronouns? *Journal of Child Language* 31, 123-152.
- Spenader, J., Smits, E. J. & Hendriks, P.  
2009 Coherent discourse solves the Pronoun Interpretation Problem. *Journal of Child Language* 36, 23-52.
- Thornton, R. & Wexler, K.  
1999 *Principle B, VP Ellipsis, and Interpretation in Child Grammar*. Cambridge, MA: MIT Press.
- Trueswell, J. C. & Gleitman, L. R.  
2007 Learning to parse and its implications for language acquisition. In: G. Gaskell (Ed.), *Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press.