

Gender Cue Effects in Children's Pronoun Processing: A Longitudinal Eye Tracking Study

Sarah Eilers, Simon P. Tiffin-Richards & Sascha Schroeder

To cite this article: Sarah Eilers, Simon P. Tiffin-Richards & Sascha Schroeder (2019) Gender Cue Effects in Children's Pronoun Processing: A Longitudinal Eye Tracking Study, *Scientific Studies of Reading*, 23:6, 509-522, DOI: [10.1080/10888438.2019.1617293](https://doi.org/10.1080/10888438.2019.1617293)

To link to this article: <https://doi.org/10.1080/10888438.2019.1617293>



© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 28 May 2019.



Submit your article to this journal [↗](#)



Article views: 266



View related articles [↗](#)



View Crossmark data [↗](#)

Gender Cue Effects in Children's Pronoun Processing: A Longitudinal Eye Tracking Study

Sarah Eilers, Simon P. Tiffin-Richards, and Sascha Schroeder

Max Planck Research Group REaD, Max-Planck-Institute for Human Development, Berlin, Germany

ABSTRACT

Children struggle with the resolution of pronouns during reading, but little is known about the sources of their difficulties. We conducted a longitudinal eye tracking experiment with 70 children in the final years of primary school. The children read sentences with a contextual resolution preference in which gender was either an informative resolution cue for the pronoun or not. We were interested in children's processing of the pronoun and their resolution preferences, as well as the effects of individual differences of Grade level and reading skill. Children's resolution ability improved with age, and good readers were more accurate than poor readers. In the eye-tracking measures, we found strong individual differences related to reading skill: Children with good reading skill took more time to read the pronoun region when pronoun gender was informative, suggesting that good readers make better use of the available information at the pronoun than poor readers.

Many beginning readers struggle with text comprehension even after having mastered fluent word reading. This suggests that word reading is necessary but not sufficient for text comprehension. Proficient readers make inferences during reading, which is one determinant of successful text comprehension (e.g., Oakhill, Berenhaus, & Cain, 2015). One example of a local inference process is pronoun resolution: Pronouns are ubiquitous in texts and easy to process by themselves as they are short and carry very little semantic meaning. On the level of word reading, pronouns are therefore not particularly challenging for beginning readers. In order to be fully understood, however, the pronoun has to be bound to an appropriate antecedent. Proficient readers routinely infer the correct antecedent using morpho-syntactic information such as gender markers (Patil, Vasishth, & Lewis, 2016). This requires the integration of information from memory across several words in a sentence or text. It has been suggested that one source of children's reading comprehension difficulties is the failure to make such inferences (Megherbi & Ehrlich, 2005; Wykes, 1981; for reviews see Nation, 2005; Oakhill, Berenhaus, & Cain, 2015; Perfetti, Landi, & Oakhill, 2005). More recently, it has been shown that children's ability to specify referents in texts accounts for unique variance in reading comprehension skill (Elbro, Oakhill, Megherbi, & Seigneuric, 2017). The ability to resolve referential relations is one of the key steps to sentence and text comprehension.

In the present longitudinal study, we examine how children at different ages and varying reading skill take different types of information into account when processing and resolving pronouns. Specifically, we investigated German children's processing and comprehension of pronouns in sentences where adults show a clear contextual resolution preference. We manipulated whether pronominal gender was an informative resolution cue or not by introducing two antecedents of either the same or a different gender.

CONTACT Sarah Eilers  eilers@posteo.de

This article has been republished with minor changes. These changes do not impact the academic content of the article. Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hssr.

© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Children's pronoun comprehension

Previous studies of children's comprehension of pronouns have produced mixed results with respect to the developmental trajectory of pronoun resolution accuracy, presumably due to differences in methods, age groups, materials and languages studied (for a review see Hickmann, Schimke, & Colonna, 2015). One study showed that children use gender information to guide online pronoun resolution during listening from 3 years of age (Arnold, Brown-Schmidt, & Trueswell, 2007), and we can assume that most children resolve pronouns correctly during listening by the time they attend primary school. However, comprehension skill moderates pronoun resolution in primary school students: In a cross-modal naming task with French 7- and 8-year-olds, Megherbi and Ehrlich (2005) demonstrated that poor comprehenders do not resolve pronouns systematically using gender information. Instead, they may resort to a default strategy where recency "overrides" other available cues.

Studies of children's reading have also shown that pronoun resolution is a source of comprehension errors. Yuill & Oakhill (1986) tested 7 to 8-year-olds' comprehension of sentences such as *On Saturday morning, Bill was going on a fishing trip with his Uncle. [...] As he carried his rod to the bus stop [...]*. Children were then asked *Who carried his rod to the bus stop?* Good comprehenders answered these questions with an error rate of 10% while poor comprehenders had an error rate of 28%. Further, Oakhill & Yuill, (1988) showed that 7- and 8-year-old children have difficulties finding the correct referent for the personal pronoun in sentences such as *Peter lent ten pence to Tom [Liz] because he [she] was very poor*. The children performed worse in the condition without an informative gender cue (16–27% error rate) compared to the condition with an informative cue (2–14% error rate). Thus, while good comprehenders performed better than poor comprehenders, both groups of children benefited from disambiguating gender information when answering the resolution question. These studies also show clearly that children struggle with the comprehension of pronouns, but they do not inform about the reading processes that are associated with resolution difficulty.

Children processing of pronouns and referential relations

Children's reading is slower and more effortful than that of adults (e.g., Gagl, Hawelka, & Wimmer, 2015). They invest extensive cognitive resources in word identification, because the translation of orthographic information into semantic representations is slower than in proficient readers. As lower-level reading requires their attention, children can invest fewer resources in higher-level processing, such as inference making and comprehension. Pronouns are very easy to process on the word level because they are both short and frequent, but they also require a higher-level integration effort, involving the retrieval of antecedent features from memory. Cue-based approaches to memory retrieval suggest that morphosyntactic cues (e.g., gender, number, grammatical case) are routinely used for resolution of pronouns (Lewis, Vasishth, & Van Dyke, 2006). Studying the use of such cues during pronoun processing can inform our understanding of how the processing demands of higher-order reading affect children of different ages and reading skill.

Pronouns have indeed been shown to be a source of difficulty in children's sentence processing, and reading ability determines pronoun processing. A self-paced reading experiment with 10-year-olds (Ehrlich, Rémond, & Tardieu, 1999) demonstrated that good comprehenders had longer reading times in clauses with a personal pronoun compared to clauses with a repeated name. In addition to reading the pronoun for a longer period of time, good comprehenders chose to press a button to display previous text more often, indicating that they adjust their rereading behavior to pronoun resolution demands. This shows that during processing, pronouns pose a specific challenge for children, arguably because they have to be resolved towards an antecedent.

Recently, eye tracking has been established as a method of choice in studying children's reading processes (for reviews see Blythe & Joseph, 2011; Schroeder, Hyönä, & Liversedge, 2015). It is favored over self-paced reading or priming methods because it allows the uninterrupted recording of multiple measures at specific points in a text.

In a pioneering eye-tracking study with 8-year-old children, Murray and Kennedy (1988) showed that good readers make more regressions in sentences that contain pronouns. Selective regressions were associated with a better comprehension of sentences with pronouns. While poor readers make more regressions in general during reading, good readers make more regressions at the pronoun than elsewhere. In a more recent eye-tracking study, Joseph, Bremner, Liversedge, and Nation (2015) examined 10-year-old children's processing of nominal anaphors. The authors compared the processing of nominal anaphors (*the vehicle*) with typical antecedents (*a truck*) and atypical antecedents (*a crane*) in stories where the anaphor was either near or distant. The authors observed more regressions when the antecedent was typical compared to when it was atypical. This finding suggests that children invest resolution effort when they are establishing a connection between anaphor and antecedent. In line with this interpretation, the authors argue that children may not resolve nominal anaphors in the distant/atypical condition at all, i.e., when resolution is most difficult. Since the study did not examine children's anaphor comprehension, however, it is still largely unclear how differences in anaphor processing are related to comprehension.

The current study

We investigated pronoun processing and comprehension in 70 German primary-school children of different reading skill in a longitudinal study. We presented sentences of the following form (see Table 1): *Paul beneidete Tessa, weil sie zu Hause einen Pool hatte* (Engl.: *Paul envied Tessa because she had a pool at home*) vs. *Paul beneidete Theo, weil er zu Hause einen Pool hatte* (Engl.: *Paul envied Theo because he had a pool at home*). We manipulated the gender of the subject and object in the main clause, resulting in sentences where pronominal gender was informative for pronoun resolution or not. In the first sentence, the gender of the pronoun is an informative resolution cue because *she* can only refer to Tessa, not to Paul. In the second sentence, gender is not informative for resolution because *he* could refer to both Paul and Theo. In the given example, however, it is plausible that Paul envied Theo because *Theo* had a pool at home. While the reading that Paul envied Theo because *Paul* had a pool at home is not strictly ruled out, it is rather laborious and less plausible. Therefore, there is a resolution preference towards Paul even in the absence of a gender cue. Similar rationales have been used in experiments with English-speaking adults (e.g., McDonald & MacWhinney, 1995; Vonk, 1984). Note that while gender-marking in German differs from English in several ways (Fagan, 2009), singular pronouns (*he/she*) are marked for natural gender as in English. The syntactic particularities of German (see the example in Table 1) do not interfere with our manipulation. In the following, we will therefore refer to our materials using English translations.

We asked a forced-choice pronoun resolution question (e.g., *Who had a pool at home?*) after every sentence to obtain resolution preference and response time (offline measures). We also recorded children's eye movements during reading (online measures). The children further completed a standardized reading comprehension test. The main research question of this study was how children of different ages and reading skill use gender and context information during pronoun processing and towards pronoun resolution.

Table 1. Structure of stimulus materials.

Gender	Item	Resolution question
Contrast	Paul beneidete Tessa , [weil sie] [zu] Hause einen Pool hatte.	Wer hatte einen Pool?
	<i>Paul envied Tessa because she had a pool at home.</i>	<i>Who had a pool?</i>
Identity	Paul beneidete Theo , [weil er] [zu] Hause einen Pool hatte.	Wer hatte einen Pool?
	<i>Paul envied Theo because he had a pool at home.</i>	<i>Who had a pool?</i>

Note. Bold text indicates the referent of the pronoun (the gender of the pronoun was counterbalanced). Square brackets indicate regions of interest for analysis. English translations are given in italics.

Comprehension of the pronoun (offline measures)

We predicted that children would answer the resolution questions more accurately after sentences that contain pronouns with an informative gender cue than no informative cue (e.g., Yuill & Oakhill, 1986). We further predicted that as children gain more reading experience with age, they should depend less on explicit gender information for resolution and instead show a more adult-like resolution preference based on the integration of sentence context. Similarly, reading skill was expected to influence resolution preferences such that better readers among the children answer the resolution questions faster and more accurately. Lastly, an interesting question concerns the relationship between reading development and individual reading skill: As children become more experienced readers, individual differences in reading skill may become less important for pronoun resolution. Such a trend would suggest that in the final years of primary school a threshold is reached such that children resolve pronouns more automatically.

Processing of the pronoun (online measures)

We analyzed reading time measures on the pronoun itself and the subsequent region. The subsequent region was taken into account to pick up effects from the pronoun that occur after it has been read. Since it is very short, effects from the pronoun may spill over onto the following word. Such a “delay” of effects has been observed in children’s syntactic processing before (Wonnacott, Joseph, Adelman, & Nation, 2016) and was shown to be developmentally relevant, as the delay reduces with age (Joseph & Liversedge, 2013). We expected to find more regressions from the pronoun region in the informative gender cue condition, i.e., when the pronoun can be resolved (Joseph et al., 2015). This would indicate that the children use the disambiguating gender information immediately for pronoun resolution. We do not expect the children to engage in resolution effort in the non-informative condition, where the pronoun can only be resolved at the end of the sentence. Besides gaze duration, we analyzed total reading time and gopast time to obtain a detailed picture of children’s rereading of the pronoun. While gaze duration is indicative of processing ease and reading fluency, total reading time and gopast time incorporate rereading following a regression. Rereading in the informative cue condition would indicate additional processing effort when disambiguating gender information is available. Longer gopast times would further indicate that children do not only regress but engage in more extensive rereading of earlier regions.

Our second research question concerned individual differences of the resolution processes in children. Children in the same Grade level differ dramatically in their individual reading ability. It is plausible that reading skill determines if and how beginning readers use gender information as a processing cue. Assuming that reading behavior at the pronoun and reading comprehension are related, we expected to see longer processing times and more regressions in the pronoun region in good readers than in poor readers. We further investigated the possibility that delayed effects occur in poor readers and therefore appear in the post-pronoun region.

Method

Participants

The children who participated in the current experiment attended two primary schools in Berlin. From the 92 original participants, we included all children who participated in both Grade 3 and Grade 4. One child was excluded because their response accuracy to comprehension questions in Grade 3 was below the chance level. The remaining 70 children completed the experiment in Grade 3, at age 8.3 years ($SD = 0.5$ years), and again 1 year later in Grade 4, aged 9.4 years ($SD = 0.5$ years). Of these 70 children, 42 were girls. All children had normal or corrected-to-normal vision.

Materials

Materials for this study comprised 24 items like the one depicted in Table 1. The study was conducted in German, but for simplicity we will illustrate the materials using English translations that leave the integrity of our stimuli intact. The sentences contained 9–12 words. Each sentence appeared in one of the two conditions (Informative Gender Cue vs. Non-Informative Gender Cue). The condition was altered by changing the names in the sentences and by adapting the pronoun accordingly. The gender of the pronoun was counterbalanced to prevent habituation effects. We took care to construct sentences with topics familiar to primary school children. For every sentence, a forced-choice resolution question was constructed from the subclause.

To support the resolution preference for the pronoun in the condition without an informative gender cue, we used implicit causality verbs that bias the resolution of the pronoun (e.g., Koornneef & Van Berkum, 2006). Only implicit causality verbs that occur in the childLex corpus, a corpus of German children's books (Schroeder, Würzner et al., 2015), were used in this experiment to ensure that the children know them. As the occurrence of these verbs in the childLex corpus is limited, the resolution preference for subject and object was counterbalanced, as were male and female pronouns. All sentences continued bias-congruent, in other words, the subordinate clause supported the preferred reading induced by the verb and there were no sentences with a conflict between verb bias and gender information. Consider the example *Clara admired Anne because she could draw so nicely*, where Anne is likely admired because she can draw nicely, or *Felix bored Pete because he always told the same stories*, where Felix is likely boring because he tells the same stories repeatedly.

To check this resolution preference, the sentences were presented to a sample of 25 adults who were recruited from local universities via mailing lists. The results from the comprehension task showed that the adults conformed to the intended resolution preference in 97% of questions, and an ANOVA with the dependent variable accuracy and the two-level factor Gender (Informative vs. Non-Informative Gender Cue) resulted in no significant effect, $F(1,48) = 2.16, p = .147$.

Children's reading skill was tested with the standardized German reading comprehension test ELFE 1–6 (Lenhard & Schneider, 2006). This test comprises three subtests targeting word, sentence and text comprehension. The raw scores for each subtest are first transformed to standardized scores and then summed up to serve as an overall indicator of children's reading skill.

Procedure

Written informed consent was collected from the children's parents ahead of the study, and oral consent was obtained from each child prior to testing. The study was approved by the ethics committee of the Max Planck Institute for Human Development, Berlin, and conforms with the Declaration of Helsinki.

Children were tested individually in a quiet room at their school during school hours. In addition, the children participated in a group session in their classroom, during which the reading comprehension test was administered. Children were tested under the same conditions in Grade 3 and Grade 4. In each session, they were assigned to one of two item lists to ensure that they read every item in only one of the cue conditions.

We used an EyeLink 1000 table-mounted eye tracker (SR Research) to record eye movements at 1000 Hz. The eye tracker was positioned under an ASUS LCD monitor (21", 120 Hz) at a 65 cm viewing distance to the child. The sentences appeared in random order, on a single line at the center of the monitor. They were presented in 14 pt Courier New using the UMass EyeTrack software (Stracuzzi & Kinsey, 2006b; version 7.10). The right eye was tracked unless tracking of the left eye considerably improved calibration. The eye tracker was calibrated using a 5-point calibration routine until calibration error reached a maximum 0.5° of visual angle. Calibration was repeated after breaks or when calibration drift were detected. After the first calibration, all children completed three practice trials. They were instructed to read the

sentences at their own pace and indicate via button press when they have finished reading. Upon pressing the button, the forced-choice pronoun resolution question appeared. To avoid confusion, the assignment of buttons to names consistently followed their position in the sentence (subject left, object right). Forty filler sentences from an unrelated experiment, including simple yes/no-comprehension questions after 25% percent of filler trials, were interspersed randomly (for details see Tiffin-Richards & Schroeder, 2015). The children answered a total of 34 comprehension questions in this experiment.

Analysis

The eye movement data were cleaned step-wise: First, each trial was inspected visually using the UMass-software EyeDoctor (Stracuzzi & Kinsey, 2006a; version 0.6.5), and y-axis drift corrections were applied to groups of fixations as necessary. Next, we applied an automatic fixation cleaning procedure as implemented in EyeDoctor. Fixations of less than 80 ms were combined with a neighboring fixation if it was within 1 character. Fixations of 40 ms or less were deleted if within 3 characters of the nearest fixation. Finally, trials with less than 5 fixations were removed (2 trials in Grade 3, 4 trials in Grade 4) and fixations under 60 ms or above 1200 ms were discarded (1.1% in Grade 3, 1.0% in Grade 4).

Four eye tracking measures were calculated for each region: *gaze duration* (sum of all fixations on a region before leaving it), *total reading time* (sum of all fixations on a region), *gopast time* (sum of all fixations from the first visit of a target region until it is left to the right), and the *probability of regression out* (likelihood that the region is exited to the left). For each measure, data points deviating more than 2.5 standard deviations from the word and subject mean were deleted (less than 2.0% of data in each group). A Pearson product-moment correlation coefficient was computed to assess the relationship of reading measures in Grade 3 and Grade 4.

We used generalized linear mixed-effects models for binomially distributed data as implemented the lme4 package (version 1.1.10; Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2016) to analyze response accuracy, and linear mixed-effects models to analyze response time and eye movement measures. Gender (Informative Gender Cue vs. Non-Informative Gender Cue) and Grade (Grade 3 vs. Grade 4) were included as effect-coded fixed effect. Reading Skill was included as a centered continuous variable. Participants and items were entered as crossed random effects in the models to allow for random intercepts for participants and items. Duration measures were log-transformed to make the distribution more normal. To ease interpretation, back-transformed model means are reported in milliseconds and probabilities, respectively. The significance of the fixed effects was determined using type-II model comparisons as implemented in the Anova function in the package car (Fox, Friendly, & Weisberg, 2013). Planned comparisons were estimated using cell-means coding and single-degree-of-freedom contrasts as implemented in the glht function in the package multcomp (Hothorn, Bretz, & Westfall, 2008).

Results

Offline measures

Resolution accuracy was positively correlated between Grade 3 and Grade 4, $r = .52$, $t(68) = 4.97$, $p < .001$, and response time was highly correlated, $r = .77$, $t(68) = 10.10$, $p < .001$. The correlation of reading skill in Grade 3 and reading skill in Grade 4 was also high, $r = .75$, $t(68) = 9.46$, $p < .001$. The model results for response accuracy and response time are summarized in Table 3, and the distributions are depicted in Figure 1.

In resolution preference, there was a main effect of Gender: As we had predicted, children were more successful in identifying the plausible antecedent in the Informative Gender condition, $M = .87$, $SE = .02$, than in the Non-Informative Gender condition, $M = .76$, $SE = .03$. Further, there was a main effect of

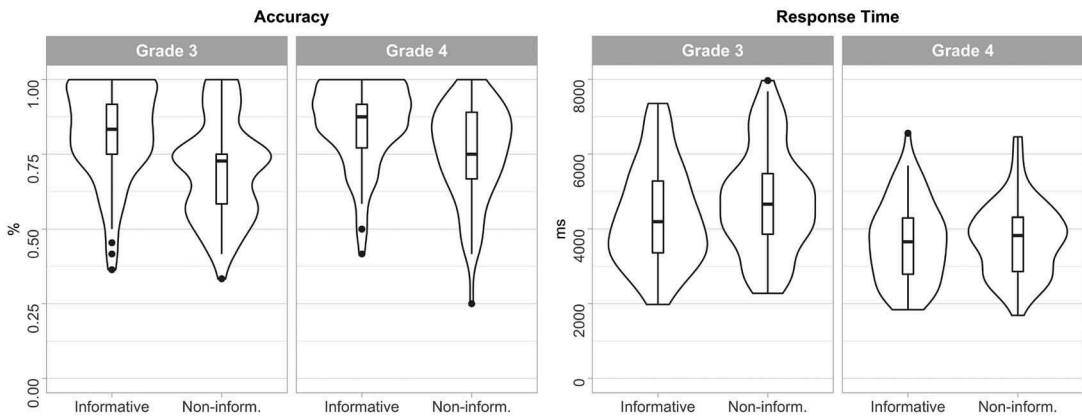


Figure 1. Distribution of resolution accuracy in percentage (panel 1) and response time in milliseconds (panel 2) for children in Grade 3 (left) and in Grade 4 (right), separately for the gender conditions. Error bars represent 2 standard errors.

Grade: Children were better at identifying the antecedent on average in Grade 4, $M = .84$, $SE = .02$, than they were in Grade 3, $M = .80$, $SE = .03$. There was also a main effect of Reading Skill on response accuracy: Good readers (1 SD above the mean), $M = .92$, $SE = .03$, were better on average than poor readers (1 SD below the mean), $M = .63$, $SE = .09$.

In response time, we found a main effect of Gender: Responses were given faster in the Informative Gender condition, $M = 3615$ ms, $SE = 128$ ms, than in the Non-Informative Gender condition, $M = 3819$ ms, $SE = 135$ ms. In addition, there was a main effect of Grade: Children were faster to respond on average in Grade 4, $M = 3275$, $SE = 117$ ms, than they were in Grade 3, $M = 4216$, $SE = 150$ ms. Finally, there was a main effect of Reading Skill: Good readers answered significantly faster, $M = 2780$ ms, $SE = 201$ ms, than poor readers, $M = 4966$ ms, $SE = 355$ ms. In addition, the Grade \times Gender interaction was significant in response time: Post-hoc analyses showed that the simple main effect of Gender was significant only in Grade 3, $t = 4.85$, $p < .001$, but not in Grade 4, $|t| < 2$, $p = .146$. The Grade \times Reading Skill interaction was also significant: Post-hoc comparisons showed that the simple main effect of Grade was smaller in good readers, $\Delta = 574$ ms, $t = 4.97$, $p < .001$, than in poor readers, $\Delta = 1412$ ms, $t = 10.44$, $p < .001$, with a significant difference effect, $t = -4.42$, $p < .001$. In summary, the effect of Gender on response accuracy remained stable with age. An unexpected effect of Gender emerged in response time, such that the informative gender cue had a facilitative effect on children's response times in Grade 3, but not in Grade 4. This may be explained by a ceiling effect such that the gender cue manipulation did not affect response times in the same way as in Grade 3.

Online measures

Gaze duration in the pronoun region in Grade 3 was positively correlated with gaze duration in Grade 4, $r = .78$, $t(68) = 10.41$, $p < .001$, as was total reading time, $r = .79$, $t(68) = 10.62$, $p < .001$ and gopast time, $r = .84$, $t(68) = 12.90$, $p < .001$. Regression probability was moderately correlated in Grade 3 and Grade 4, $r = .45$, $t(68) = 4.20$, $p < .001$.

Descriptive statistics for eye tracking measures in the pronoun and post-pronoun region are given in Table 2, and the results from the mixed-effect models are given in Table 4. To describe the effect of individual differences in reading skill on the eye movement measures, we quantified the effect of reading skill at 1 SD above and 1 SD below the mean reading score using contrasts.

Table 2. Overview of observed means.

	Pronoun region		Post-pronoun region	
	Inf. Gender	Non-Inf. Gender	Inf. Gender	Non-Inf. Gender
	<i>Gaze duration</i>			
Grade 3	569 (376)	557 (375)	364 (245)	364 (223)
Grade 4	455 (291)	453 (279)	305 (191)	299 (158)
	<i>Total reading time</i>			
Grade 3	819 (552)	795 (567)	516 (378)	516 (356)
Grade 4	662 (450)	614 (427)	428 (315)	401 (273)
	<i>Gopast time</i>			
Grade 3	761 (609)	761 (713)	495 (435)	482 (402)
Grade 4	643 (630)	595 (523)	413 (365)	377 (303)
	<i>Regression probability</i>			
Grade 3	.20 (.02)	.20 (.02)	.18 (.02)	.17 (.02)
Grade 4	.20 (.02)	.15 (.02)	.20 (.02)	.17 (.02)

Note. Standard deviations are given in parentheses. Inf. Gender = informative gender, Non-Inf. Gender = non-informative gender.

Table 3. Results from the linear mixed-effects models over offline measures. χ^2 -values for response accuracy. ANOVA *F*-values for response time.

	Offline measures	
	Accuracy	Time
Gender	66.76 ***	19.11 ***
Grade	20.37 ***	243.39 ***
Reading	8.92 **	26.78 ***
Gender × Grade	0.33	4.05 *
Gender × Reading	0.59	0.90
Grade × Reading	0.07	6.52 *
Gender × Grade × Reading	2.47	0.01

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Table 4. Results from the linear mixed-effects models over online measures in the pronoun region. ANOVA *F*-values for reading time measures. χ^2 -values for probability of regression out.

	Gaze	Total	Gopast	Regr. prob.
	<i>Pronoun region</i>			
Gender	0.08	9.08 **	2.35	4.31 *
Grade	139.76 ***	190.77 ***	123.76 ***	3.25
Reading	23.24 ***	40.03 ***	21.66 ***	2.36
Gender × Grade	0.01	1.35	1.24	4.70 *
Gender × Reading	1.58	3.85 *	5.95 *	6.27 *
Grade × Reading	22.54 ***	3.05	6.90 **	0.15
Gender × Grade × Reading	3.60	0.00	0.02	0.26
	<i>Post-pronoun region</i>			
Gender	0.04	0.76	2.69	3.63
Grade	120.89 ***	147.00 ***	101.28 ***	0.52
Reading	36.47 ***	46.76 ***	23.51 ***	0.08
Gender × Grade	0.03	0.26	0.00	0.67
Gender × Reading	1.17	1.32	2.01	0.02
Grade × Reading	33.78 ***	13.10 ***	6.89 **	1.56
Gender × Grade × Reading	0.12	0.03	1.67	0.50

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$. Regr. prob. = Regression probability.

Pronoun region

In gaze duration, there was a main effect of Grade: Children in Grade 4 showed shorter gaze durations than in Grade 3 ($\Delta = 58$ ms). There was also a main effect of Reading Skill: Good readers had significantly shorter gaze durations in the pronoun region than poor readers ($\Delta = 178$ ms). In addition,

the Grade \times Reading Skill interaction was significant. The simple main effect of Grade was not significant in the good readers, $|t| < 1$, $p = .476$, but was significant in the poor readers, $\Delta = 146$ ms, $t = 10.83$, $p < .001$. Neither the main effects of Gender nor any interaction involving Gender were significant.

In total reading time, there were main effects of Gender and Grade: Total reading time was higher in the Informative condition, $M = 608$, $SE = 24$ ms, than in the Non-Informative condition, $M = 577$ ms, $SE = 23$ ms. The main effect of Grade showed that children became faster readers in Grade 4 ($\Delta = 107$ ms). There was also a main effect of Reading Skill: Good readers spend less time in the pronoun region than poor readers ($\Delta = 675$ ms). In addition, the Gender \times Reading Skill interaction was significant: The simple main effect of Gender was significant in good readers, $\Delta = 59$ ms, $t = 2.89$, $p < .01$, but not poor readers, $|t| < 1$, $p = .562$. The means of good (+1 SD) and poor (-1 SD) readers in the two gender cue conditions are depicted in Figure 2 (left panel). In addition, the full distribution of total reading times in the two gender cue conditions as a function of reading skill is provided in Figure 3 (left panel).

In gopast time, we found no main effect of Gender but a main effect of Grade: Children had shorter gopast times in Grade 4, $M = 503$ ms, $SE = 20$ ms, than in Grade 3, $M = 697$ ms, $SE = 28$ ms. In addition, there was a main effect of Reading Skill and interactions of Gender \times Reading Skill (see Figures 2 and 3, mid panel). The simple main effect of Gender was significant in good readers, $\Delta = 59$ ms, $t = 2.70$, $p < .01$, but not poor readers, $|t| < 2$, $p = .118$. There was also an interaction of Grade \times Reading Skill. From Grade 3 to Grade 4, good readers significantly reduced their gopast times, $\Delta = 70$ ms, $t = 2.79$, $p < .01$. In poor readers, this reduction was significantly larger, $\Delta = 269$ ms, $t = 8.06$, $p < .001$, $t = -2.63$, $p < .05$.

In regression probability, there was a main effect of Gender, as well as a significant Gender \times Reading Skill interaction. In addition, the interaction of Gender \times Reading Skill was significant (Figures 2 and 3, right panel). The simple main effect of Gender was only significant in good readers, $\Delta = .08$, $t = 2.87$, $p < .01$, but not in poor readers, $|t| < 2$, $p = .187$. Good readers made more regressions when Gender was informative, $M = .20$, $SE = .03$, than when it was not informative, $M = .12$, $SE = .02$. In addition, the Gender \times Grade interaction was also significant: The simple main effect of Gender was significant in Grade 4, $t = 2.74$, $p < .01$, but not in Grade 3, $|t| < 1$, $p = .884$. In summary, our findings suggest that children with better reading skill spend

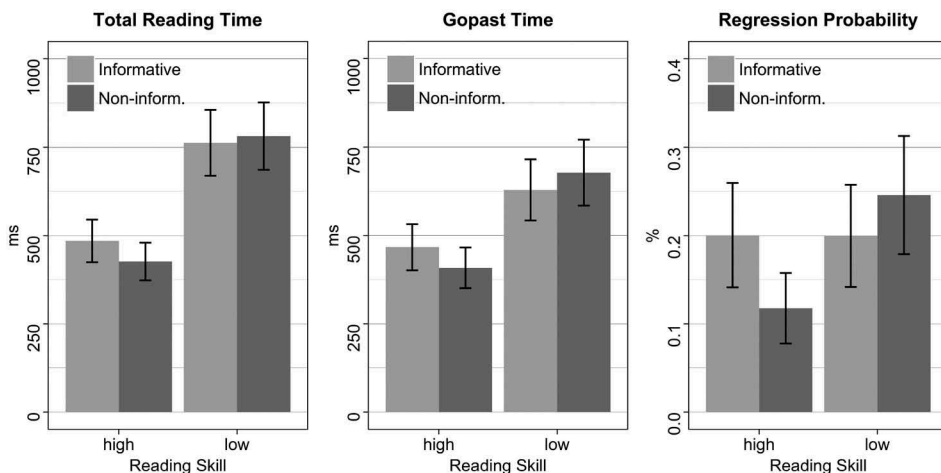


Figure 2. Means for total reading time (left panel), gopast time (mid panel), and regression probability (right panel), back-transformed to milliseconds and probability, respectively, at 1 SD above the mean (good reading skill) and 1 SD below the mean (poor reading skill), in the two Gender conditions. Error bars represent 2 standard errors.

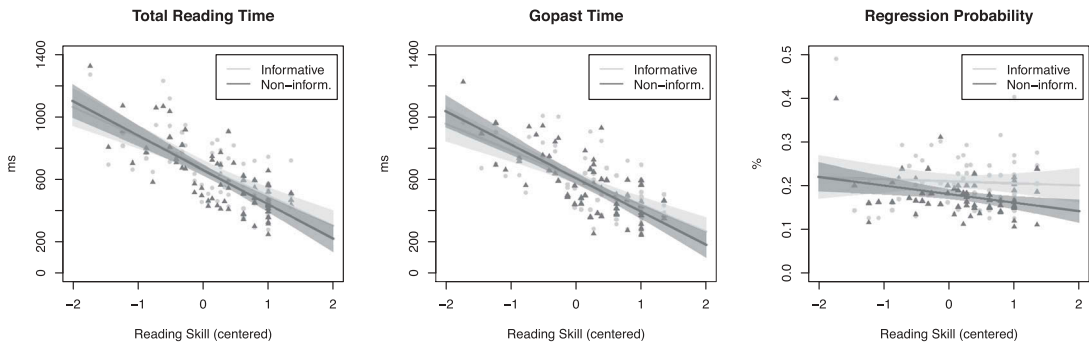


Figure 3. Distribution of total reading time (left panel), gopast time (mid panel), and regression probability (right panel) in the two Gender conditions as a function of reading skill (centered), back-transformed to milliseconds and probability, respectively. Confidence intervals represent two standard errors.

more processing time in the pronoun region during the second pass when it contains useful information for resolution. Against our expectations, these effects do not change from Grade 3 to Grade 4.

Post-pronoun region

In the post-pronoun region, we found no significant effects of Gender or interactions with Gender in any of the reported measures (see Table 4). There were however main effects of Grade and Reading Skill, as well as interactions of Grade \times Reading Skill in gaze duration, total reading time and gopast time: In gaze duration the simple main effect of Grade was not significant for good readers, $|t| < 1$, $p = .476$, but was significant for poor readers, $\Delta = 164$ ms, $t = 10.83$, $p < .001$. Similarly, in total reading time, the simple main effect of Grade was not significant for good readers, $|t| < 2$, $p = .072$, but was significant for the poor readers, $\Delta = 229$ ms, $t = 9.78$, $p < .001$. In gopast time, the simple main effect of Grade was significant for good readers, $\Delta = 34$ ms, $t = 2.07$, $p < .05$, as well as the poor readers, $\Delta = 157$ ms, $t = 7.32$, $p < .001$. The difference of the effect of Grade for good and poor readers was also significant, $t = 2.63$, $p < .01$. In regression probability, we found no effects at all in the post-pronoun region.

In summary, as there were no effects of Gender, or interactions of Gender \times Grade or Gender \times Reading Skill in the post-pronoun region, we may conclude that there were no spill-over effects of Gender information from the pronoun region. The effects of Grade are similar to those found in the pronoun region, indicating that the children become faster, more fluent readers with age.

Exploratory analyses of antecedent position

To further explore how children resolved the pronoun in our experiment, we conducted an additional analysis of the effects of antecedent position. Recall that the resolution preference was counterbalanced in the sentences. In each Gender condition therefore half of the antecedents were in subject position and therefore mentioned first, while the other half were in object position and mentioned second. We calculated a set of additional models in which we added the factor Antecedent Position (Mentioned First vs. Mentioned Second), everything else being equal.

The effects of Gender, Age and Reading Skill reported above remained significant in the offline measures after the addition of Antecedent Position into the model. There was a main effect of Position on resolution ability: Children were better at selecting the plausible antecedent in the Mentioned Second condition (object antecedent), $M = .75$, $SE = .02$, than in the Mentioned First condition (subject antecedent), $M = .67$, $SE = .02$, $t = 31.45$, $p < .001$. In response time, there was also a main effect of Antecedent Position, such that

questions in the Mentioned Second condition were answered significantly faster, $M = 3495$ ms, $SE = 132$ ms, than in the Mentioned First condition, $M = 3950$ ms, $SE = 150$ ms, $t = 9.48$, $p < .01$. In both measures, there were no interactions with other factors, all $t < 2.2$. In the online measures, i.e., gaze duration, total reading time, gopast time and regression probability, there were no effects of Antecedent Position, all $ts < 2$. To summarize these results: Children are faster and more accurate, i.e., conform to the plausible context more often, when the pronoun refers to the second-mentioned, or last-mentioned antecedent. However, effects of Gender, Age and Reading Skill remain robust after accounting for Antecedent Position. We saw no indication that Position influences online reading behavior at the pronoun.

Discussion

The present study investigated how children use gender and context information when resolving pronouns. We presented sentences containing pronouns with informative and non-informative gender cues to children in Grade 3 and again in Grade 4. We found that disambiguating gender information had a positive effect on children's ability to determine the correct referent after reading. While children's general resolution ability improved from Grade 3 to Grade 4, the effect of gender information on resolution remained stable. We further showed that disambiguating gender information on the pronoun affected late processing measures in children, but this effect was moderated by reading skill: Only children with high reading skill used the gender information immediately during reading, such that they invest more processing time when an informative gender cue can be used to resolve the pronoun on the spot. We conclude that children with higher reading skill invest available processing resources towards local inference generation. We discuss the findings from the offline and online measures separately in the remainder of the Discussion.

Comprehension of the pronoun

Results from our offline measures showed that children clearly struggled with the assignment of an antecedent for the pronoun in our study, particularly in the absence of gender information as a resolution cue. When the pronoun could only be resolved on account of context information, given by the main verb and the subclause, children's accuracy dropped significantly. This is in line with earlier observations (Yuill & Oakhill, 1986). We interpret these findings to suggest that German children in Grade 3 and Grade 4 typically need explicit resolution cues for the resolution of pronouns when reading. Although they allowed themselves more response time when there was no informative gender cue, many children seem unable to find a plausible antecedent for the pronoun using context information alone. Further, while children's overall resolution ability improved with age, the effect of the gender information cue on resolution remained stable. This indicates that children in the final phase of primary school may not have developed the necessary inference skills to resolve pronouns in the absence of explicit cues, regardless of Grade level or reading skill.

But how do children decide on an antecedent? The results from our additional analysis of antecedent position showed that children often chose the last-mentioned person as the antecedent for the pronoun, even when this interpretation is not supported by the sentence context. This might indicate that children resort to default resolution strategies, as has previously been shown in listening studies (e.g., Megherbi & Ehrlich, 2005). Further, the study of children's comprehension of relative clauses has shown that children predominantly interpret object relative clauses as subject relative clauses (e.g., Adani, Van der Lely, Forgiarini, & Guasti, 2010). The authors suggest that the children fail to interpret the syntactic dependencies. This supports our interpretation that children do not sufficiently take the sentence context into account when resolving the pronoun. Since this additional analysis is based on exploratory results, it should be treated with some caution. Our results, however, certainly warrant further investigation into children's strategies of pronoun resolution during reading, and their effects on sentence and text comprehension.

Online processing of the pronoun region

In addition to the offline comprehension measures, we recorded children's eye movements in the pronoun area to obtain a detailed picture of the incremental reading processes at the pronoun. While the offline measures reflect children's response behavior after having read the whole sentence, the online measures provide information on the moment-to-moment processing of the pronoun when it is encountered. We were interested in the processing of the pronoun region because it indicates whether the children use information from the gender cue immediately during reading. The results are clear-cut. First, we found that when children initiate regressions and rereading, they did so directly from the pronoun region and not the post-pronoun region. This is true even for the poor readers, who did not show the delayed effects of processing we had hypothesized. Our results are consistent with Joseph et al. (2015), who found evidence of anaphoric processing in children beginning directly on the anaphor itself. Second, effects of gender information only occurred in the good readers, and only in late processing measures, specifically regression probability and total reading time. Because there were no effects in gaze duration, the effects in total reading time are entirely attributable to rereading. Based on the results for gopast times, we can say that the informative pronoun does not induce extensive rereading of the previous sentence regions. This indicates that good readers, but not poor readers, adjust rereading time of the pronoun region. The individual differences in online processing were substantial: Only children with good reading skill had longer total reading times in the pronoun region and made more regressions from the pronoun in the informative cue condition, when the antecedent was unambiguous. Our results are compatible with cue-based approaches to memory retrieval in sentence processing (e.g., Lewis, Vasishth, & Van Dyke, 2006; Patil et al., 2016), which assume that proficient readers use different types of information, including morpho-syntactic gender information, towards pronoun resolution immediately when it becomes available. It appears that when reading processing is effortful for children, they may not allocate attention to these retrieval cues. Another explanation is that children lack the necessary reading experience to identify morpho-syntactic information, such as pronoun gender, as a relevant cue during online reading. In both scenarios, beginning readers may then resort to default strategies for pronoun resolution.

Good readers among the children use gender information immediately when it is informative to resolve the pronoun, hence their longer processing times. This suggests that children with good reading comprehension skill process key areas in a sentence differently from children with poor reading comprehension skill: Children with good comprehension skill reread selectively and adjust their processing time to the informative content of the pronoun. This is in line with earlier findings for individual differences in children's regression behavior (Murray & Kennedy, 1988).

Studies of children's reading development have repeatedly found that faster word decoding does not necessarily lead to successful comprehension (for a review see Nation, 2005). It is noteworthy that although children's reading fluency improved considerably from Grade 3 to Grade 4, the effect of reading skill on the use of the disambiguating gender information remained stable in our study. Thus, despite faster word reading, children with poor reading skill did not automatically "catch up" in their pronoun comprehension or the way in which they process the pronoun region during reading.

Considering our offline and online results together, we conclude that many children are unable to resolve a pronoun during sentence reading when they cannot do so immediately. In other words, when the children cannot resolve the pronoun on the spot based on an explicit, informative gender cue, they are unlikely to do so later in the sentence or after reading. It seems that when resolution is difficult because it requires integration of context information, many children do not invest the necessary effort to construct a coherent representation of what they have read.

In sum, the results of our study show that German children at the end of primary school still struggle with the resolution of pronouns during reading, particularly when they need to take the sentence context into account to identify a plausible antecedent. While the accuracy of pronoun

resolution generally improved from Grade 3 to Grade 4, children in Grade 4 still benefit from an explicit, informative gender cue and have not yet reached adult resolution efficiency.

Acknowledgments

Simon P. Tiffin-Richards is now at Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin.

Sascha Schroeder is now at the University of Göttingen, Department of Psychology, Waldweg 26, 37073 Göttingen.

Conflict of interest

The authors declare no conflict of interest.

Ethics statement

The study was approved by the ethics committee of the Max Planck Institute for Human Development, Berlin, and conforms with the Declaration of Helsinki.

References

- Adani, F., Van der Lely, H. K., Forgiarini, M., & Guasti, M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua*, *120*(9), 2148–2166. doi:10.1016/j.lingua.2010.03.018
- Arnold, J. E., 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. doi:10.1080/01690960600845950
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01
- Blythe, H. I., & Joseph, H. (2011). Children's eye movements during reading. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *Oxford Handbook on Eye Movements* (pp. 643–662). Oxford, UK: Oxford University Press.
- Ehrlich, M. F., Rémond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing*, *11*(1), 29–63. doi:10.1023/A:1007996502372
- Elbro, C., Oakhill, J., Megherbi, H., & Seigneuric, A. (2017). Aspects of pronominal resolution as markers of reading comprehension: The role of antecedent variability. *Reading and Writing*, *30*(4), 813–827. doi:10.1007/s11145-016-9702-4
- Fagan, S. M. (2009). *German: A linguistic introduction*. Cambridge, UK: Cambridge University Press.
- Fox, J., Friendly, M., & Weisberg, S. (2013). Hypothesis tests for multivariate linear models using the car package. *The R Journal*, *5*(1), 39–52. doi:10.32614/RJ-2013-004
- Gagl, B., Hawelka, S., & Wimmer, H. (2015). On sources of the word length effect in young readers. *Scientific Studies of Reading*, *19*(4), 289–306. doi:10.1080/10888438.2015.1026969
- Hickmann, M., Schimke, S., & Colonna, S. (2015). From early to late mastery of reference. In L. Serratrice & S. E. M. Allen (Eds.), *The Acquisition of Reference* (pp. 181–211). Amsterdam, Netherlands: John Benjamins.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363. doi:10.1002/(ISSN)1521-4036
- Joseph, H. S., Bremner, G., Liversedge, S. P., & Nation, K. (2015). Working memory, reading ability and the effects of distance and typicality on anaphor resolution in children. *Journal of Cognitive Psychology*, *27*(5), 622–639. doi:10.1080/20445911.2015.1005095
- Joseph, H. S., & Liversedge, S. P. (2013). Children's and adults' on-line processing of syntactically ambiguous sentences during reading. *PLoS One*, *8*(1), e54141. doi:10.1371/journal.pone.0054141
- Koornneef, A. W., & Van Berkum, J. J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, *54*(4), 445–465. doi:10.1016/j.jml.2005.12.003
- Lenhard, W., & Schneider, W. (2006). *ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen, Germany: Hogrefe.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*, 447–454. doi:10.1016/j.tics.2006.08.007
- McDonald, J. L., & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, *34*, 543. doi:10.1006/jmla.1995.1025

- Megherbi, H., & Ehrlich, M. F. (2005). Language impairment in less skilled comprehenders: The on-line processing of anaphoric pronouns in a listening situation. *Reading and Writing, 18*(7), 715–753. doi:10.1007/s11145-005-8131-6
- Murray, W. S., & Kennedy, A. (1988). Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analyses. *The Quarterly Journal of Experimental Psychology, 40*(4), 693–718. doi:10.1080/14640748808402294
- Nation, K. (2005). Children's reading comprehension difficulties. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 248–265). Oxford, UK: Blackwell.
- Oakhill, J., & Yuill, N. (1988). Pronoun resolution in skilled and less skilled comprehenders: Effects of memory load and inferential complexity. *Language and Speech, 29*(1), 25–37. doi:10.1177/002383098602900104
- Oakhill, J. V., Berenhaus, M. S., & Cain, K. (2015). Children's reading comprehension and comprehension difficulties. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 344–359). Oxford, UK: Oxford University Press.
- Patil, U., Vasissth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology, 7*, 329. doi:10.3389/fpsyg.2016.00329
- Perfetti, C., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford, UK: Blackwell.
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Schroeder, S., Hyönä, J., & Liversedge, S. P. (2015). Developmental eye-tracking research in reading: Introduction to the special issue. *Journal of Cognitive Psychology, 27*, 500–510. doi:10.1080/20445911.2015.1046877
- Schroeder, S., Würzner, K. M., Heister, J., Geyken, A., & Kliegl, R. (2015). childLex: A lexical database of German read by children. *Behavior Research Methods, 47*(4), 1085–1094. doi:10.3758/s13428-014-0528-1
- Stracuzzi, D., & Kinsey, J. (2006a). *EyeDoctor*, (Version 0.6.5) [Computer software]. Retrieved from <http://www.psych.umass.edu/eyelab/software/>
- Stracuzzi, D., & Kinsey, J. (2006b). *EyeTrack*(Version 7.10m) [Computer software]. Retrieved from <http://www.psych.umass.edu/eyelab/software/>
- Tiffin-Richards, S. P., & Schroeder, S. (2015). Children's and adults' parafoveal processes in German: Phonological and orthographic effects. *Journal of Cognitive Psychology, 27*(5), 531–548. doi:10.1080/20445911.2014.999076
- Vonk, W. (1984). Eye Movements during the comprehension of pronouns. In A. G. Gale & F. Johnson (Eds.), *Theoretical and Applied Aspects of Eye Movement Research* (pp. 203–212). Amsterdam, Netherlands: North-Holland.
- Wonnacott, E., Joseph, H. S., Adelman, J. S., & Nation, K. (2016). Is children's reading "good enough"? Links between online processing and comprehension as children read syntactically ambiguous sentences. *The Quarterly Journal of Experimental Psychology, 69*(5), 855–879. doi:10.1080/17470218.2015.1011176
- Wykes, T. (1981). Inference and children's comprehension of pronouns. *Journal of Experimental Child Psychology, 32* (2), 264–278. doi:10.1016/0022-0965(81)90080-1
- Yuill, N., & Oakhill, J. (1986). Understanding of anaphoric relations in skilled and less skilled comprehenders. *British Journal of Psychology, 79*(2), 173–186. doi:10.1111/j.2044-8295.1988.tb02282.x