Applying the Visual World Paradigm in the Investigation of Preschoolers' Online Reference Processing in a Continuous Discourse

Abigail Toth[1,2] **(a.g.toth@rug.nl)**
[1] Department of Linguistics, University of Alberta
4-32 Assiniboia Hall, Edmonton, AB, T6G 2E7, Canada
[2] Department of Artificial Intelligence, University of Groningen
Nijenborgh 9, 9747 AG Groningen, The Netherlands

Monique Charest (mcharest@ualberta.ca)
Department of Communication Sciences and Disorders, University of Alberta
2-70 Corbett Hall, Edmonton, AB, T6G 2G4, Canada

Jacolien van Rij (j.c.van.rij@rug.nl)
Department of Artificial Intelligence, University of Groningen
Nijenborgh 9, 9747 AG Groningen, The Netherlands

Juhani Järvikivi (jarvikivi@ualberta.ca)
Department of Linguistics, University of Alberta
4-32 Assiniboia Hall, Edmonton, AB, T6G 2E7, Canada

Abstract

Using a novel adaptation of the visual world eye-tracking paradigm we investigated children's and adults' online processing of reference in a naturalistic language context. Participants listened to a 5-minute long storybook while wearing eye-tracking glasses. The gaze data were analyzed relative to the onset of referring expressions (i.e., full noun phrases (NPs) and pronouns) that were mentioned throughout the story. We found that following the mention of a referring expression there was an increase in the proportion of looks to the intended referent for both children and adults. However, this effect was only found early on in the story. As the story progressed, the likelihood that participants directed their eye gaze towards the intended referent decreased. We also found differences in the eye gaze patterns between NPs and pronouns, as well as between children and adults. Overall these findings demonstrate that the mapping between linguistic input and corresponding eye movements is heavily influenced by discourse context.

Keywords: visual world paradigm; eye-tracking; reference processing; discourse

Introduction

During spoken communication, we use different types of referring expressions in order to specify people, places and things. These include both full noun phrases (NPs) (e.g., 'Sarah', 'the bear') and pronouns (e.g., 'she', 'it'). In order for communication to be successful, speakers must choose appropriate referring expressions and listeners must rapidly map those referring expressions onto the intended referents. One method that has been used to investigate the online comprehension of reference is the visual world eye-tracking paradigm (VWP) (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In the VWP, an individual's eye movements are monitored as they receive spoken language input and view a visual scene. The eye gaze response relative to the spoken language input is taken to reflect underlying processes involved in online language comprehension.

In a seminal paper, Cooper (1974) found that when people were simultaneously presented with spoken language input and a visual scene, they naturally directed their eye gaze towards entities in the visual scene that are semantically related to the language being heard. For example, when participants heard phrases such as 'suddenly I noticed a hungry lion' they fixated on a lion present in the visual scene ~200 milliseconds (ms) after hearing 'lion'. Cooper proposed that the relationship between spoken language input and eye gaze fixations could be viewed as an active online process, and as such could be used to investigate online language processing. Tanenhaus and colleagues (1995) further investigated the influence of the visual scene on spoken language comprehension. They recorded participants' eye movements as participants followed spoken instructions and manipulated real objects visible in front of them. They found that participants fixated on target objects ~250 ms after hearing the word that uniquely identified the target. For example, when participants heard instructions such as 'touch the starred yellow square', they fixated on the target 250 ms after hearing 'starred' when there was only one starred object. However, if participants were given the same instruction and there were two starred objects, they did not fixate on the target until 250 ms after hearing 'yellow', highlighting the high temporal resolution between linguistic input and corresponding eye

movements. Both these studies were instrumental in the development of the VWP as a tool for investigating online language processing.

The mapping between linguistic input and corresponding eye movements has also been utilized to investigate language processing that relies on inference, as is the case in online pronoun resolution (e.g., Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Järvikivi, Van Gompel, Hyönä, & Bertram, 2005). Typically in these studies, participants view scenes with two (or more) referents while listening to passages that contain an ambiguous pronoun. Because pronouns do not have a fixed meaning, there is not a direct association between the linguistic input (i.e., 'he') and a corresponding referent in the visual scene. Thus, listeners must infer which referent the pronoun refers to and eye gaze is taken to reflect this process. For example, VWP studies have reported that following the mention of an ambiguous third person singular pronoun (i.e., 'he'), there is an increase in the proportion of looks to the grammatical subject of the preceding sentence/clause (e.g., Järvikivi, et al., 2005; Kaiser & Trueswell, 2008). This increase in the proportion of looks has been taken to suggest that participants interpreted the pronoun as co-referring with the preceding subject, providing further evidence for what is known as the subject bias (e.g., Crawley, Stevenson, & Kleinman, 1990; Frederiksen, 1981).

In addition to high temporal resolution, another advantage of the VWP is that it does not require participants to read or carry out demanding tasks, and therefore can be used to investigate online language processing in young children. This allows for direct comparisons between children and adults without the potential confounds introduced by response requirements. For example, VWP studies have reported that children as young as 2.5 to 4 years old appear to be sensitive to the subject bias (e.g., Hartshorne, Nappa, & Snedeker, 2015, for an overview; Järvikivi, Pyykkönen-Klauck, Schimke, & Hemforth, 2014; Pyykkönen, Matthews, & Järvikivi 2010; Song & Fisher 2005; 2007). However, the increase in the proportion of looks to the grammatical subject usually did not occur until relatively late (i.e., 1200 ms) after the pronoun onset, suggesting there is still a difference between children and adults.

Another advantage that has been attributed to the VWP is that it can be used to investigate language processing under relatively realistic conditions. This is primarily because the comprehension processes can proceed uninterrupted by response requirements. However, the majority of previous VWP studies have used carefully designed tasks that often encourage participants to carefully look at the visual scene. Furthermore, participants were usually presented with a series of isolated experimental items, where each item was no more than 2-3 sentences, and thus lacked any sort of rich context. This means that each item introduced a new situation or topic, for which participants had no context. In sum, previous applications of the VWP may not accurately reflect naturalistic language processing.

To date, only a single study has used the VWP to investigate reference processing in a continuous discourse. Engelen, Bouwmeester, de Brain and Zwaan (2014) had children listen to a 7-minute long story while viewing a display containing four black and white animal line drawings. They analyzed eye gaze data for both full NPs (e.g., 'rabbit') and pronouns (e.g., 'he') and found that following the onset of a referring expression there was an increase in proportion of looks to the target. However, they also found that overall target fixations (following an NP or pronoun) decreased as the story unfolded over time. However, given that participants viewed the same simple display for the entire 7-minutes, it is possible that the overall decrease was an artifact of fatigue or boredom. To date no study has used the VWP to investigate reference processing in a context where both the language input and visual scene reflect a natural language setting.

Present Study

The present study applied a novel adaptation of the visual world eye-tracking paradigm (VWP) in order to explore online reference processing in a naturalistic language setting. Children and adults listened to a 5-minute story containing multiple animal characters while wearing eye-tracking glasses (ETG). We opted to use ETG over a more traditional table-mounted eye tracker because we wanted to keep the language processing context as naturalistic as possible. The ETG are akin to normal reading glasses and allow for participants to move more freely throughout the duration of the experiment. We analyzed eye gaze patterns with respect to the onset of referring expressions (full NPs and pronouns) mentioned throughout the story. Given the novelty of the methodological application it was important to be able to compare the eye gaze patterns between the two types of referring expressions, as well as between children and adults. Our primary goal was to explore language mediated eye movements outside the context of a carefully designed VWP task. We were interested in what eye gaze patterns could tell us about the processing of continuous discourse in a naturalistic language setting.

Method

Participants

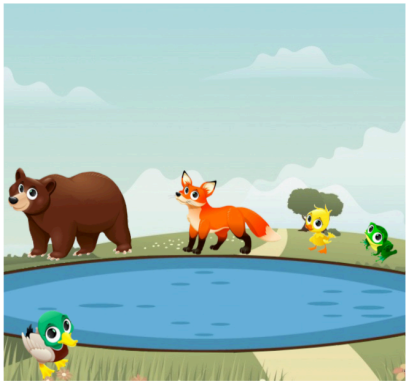
Thirty-five native English-speaking children recruited from preschools and daycares in Edmonton, Alberta, participated in the study. Written parental consent was obtained prior to participation and children received stickers and a t-shirt in exchange for their participation. Despite the fact that the ETG were supposed to be child-friendly, during the analysis it became evident that there

was large amount of gaze loss across participants. This was because the angle between the cameras on the ETG and children's pupils was too large, meaning that the ETG were too big to accurately keep track of eye gaze for some children. This resulted in 17 children being excluded from the analysis. An additional 3 children were excluded because they did not fit the age range (> 6 years old). This resulted in 15 children (7 female; mean age = 4.8 years; range 4.2-5.6) being included in the final analysis. All children had normal vision and hearing based on parental-report.

Sixteen native English speaking adults also participated in the study to serve as a control group. All adults were undergraduate students from the University of Alberta and received partial course credit for their participation. Written consent was obtained prior to participation. Four adults were excluded from the analysis due to technical issues with the ETG. This resulted in 12 adults (10 female; mean age = 20 years; range 18.2-22) being included in the final analysis. All adults had normal vision and hearing based on self-report.

Materials

A 22-page electronic storybook was constructed to be similar in style to an everyday storybook that would be read to children. The story was about a group of animal friends helping a duckling find his father. It contained multiple referring expressions in the linguistic discourse, with corresponding referents in the illustrations. The story began with a single character and after every 3-5 pages a new character was introduced so that it ended with 5 characters in total. All characters were referred to using the masculine pronoun 'he' to ensure ambiguity. The storybook illustrations were created using images from freepik.com. The audio was recorded by a female native speaker of English in a sound-attenuated booth. The illustrations and audio were then pieced together into a 5 minute and 26 second long .mp4 video, where the pages flipped as if it were a real book. An example illustration and associated dialogue can be seen Figure 1 below.



'But before anyone could start looking, Duckling spotted Daddy Duck across the pond! He flapped his wings with excitement. Daddy Duck looked up and saw Duckling. He sighed with relief and started swimming across the pond.'

Figure 1: Example illustration and associated dialogue

Critical Items

Thirty-six full NPs (i.e., character names) and 10 ambiguous pronouns (i.e., he) were selected as experimental items. These items were selected with the criteria that they did not overlap with other referring expressions, meaning that another referring expression could not occur within the ~1200 ms window following their onset. Furthermore, pronouns had to follow a clause where both the grammatical subject and object were animal characters. Because this was a natural story, there was variation in the input that both preceded and followed the critical pronouns. However, it should be noted that at pronoun onset (and for a period of time afterwards), all critical pronouns were ambiguous. Two examples can be seen below and the full set can be found in the supplementary materials[1].

1) Fox thanked Bear. He wanted to play a different game.
2) Bear told Fox to go and hide. He started counting to five.

Procedure

All children were tested individually at their preschool or daycare. The children sat approximately 50 cm in front of a Lenovo laptop and were first familiarized to the animal characters by being shown each animal individually and then being asked to name the animal. In the event that the child misnamed the animal they were corrected. The children were then told they would listen to a short story about the animals while wearing special ETG. The ETG were placed on the child's head and secured using an adjustable strap. The children completed a 3-point calibration and then listened to the electronic storybook. The eye gaze data were collected with SensoMotoric Instruments (SMI) ETG wireless 2 eye-tracking glasses, which included a built-in high-definition scene camera that recorded all audio and video. Registration was binocular with a sampling rate of 60 Hz (16.6ms/frame). After listening to the storybook children were asked a series of five comprehension questions in order to ensure that they had been paying attention. Children had to answer at least four questions correctly in order to be included in the analysis. All children met this criterion.

All adults were tested at The Centre for Comparative Psycholinguistics at the University of Alberta following a similar procedure.

Gaze coding

The gaze data were coded frame-by-frame using Noldus ObserverXT software (Noldus Information Technology, 2012). The areas of interest were each of the five animal characters. Eye gaze that fell outside of interest areas (IAs) was coded as 'elsewhere' or in the case of trackloss was coded as 'NA'. This resulted in a data frame in which each IA had a separate column and every row represented a single frame (~16.7 ms of time). For each frame the IA columns either had a value of 0 (gaze not within IA) or 1

[1]The supplementary materials can be found at: https://git.lwp.rug.nl/a.g.toth/VWP- discourse

(gaze within IA). The gaze data were then binned into 5-frame time bins each representing ~83 ms so that the values in the IA columns could have values between 0 and 5.We were interested in analyzing looks to the target referent (versus looks elsewhere), which in the case of pronouns was coded as the subject of the preceding clause. The time window we were interested in was 2 bins before the referring expression onset up until 14 bins after the referring expression onset (~1415 ms in total).

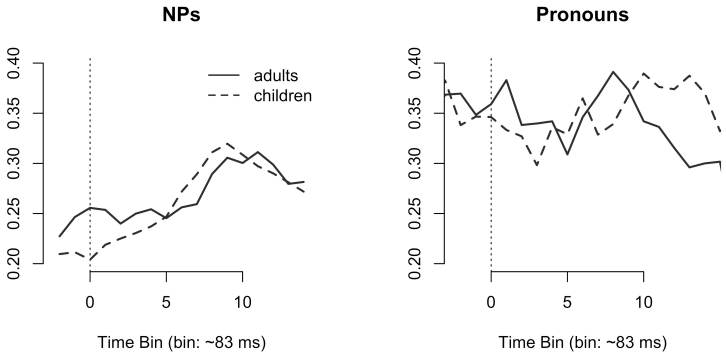


Figure 2: Average proportion of target looks across the time bin analysis window (~1415 ms)

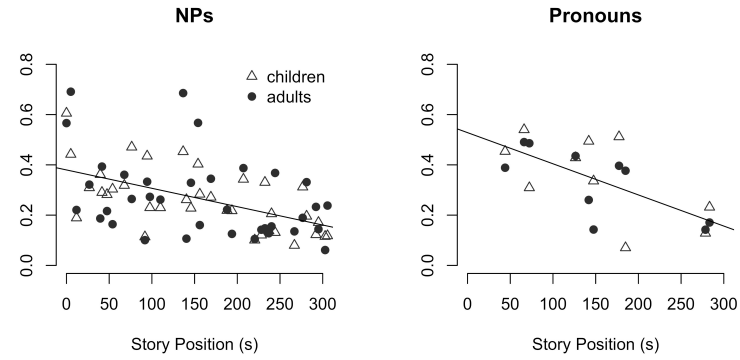


Figure 3: Average proportion of target looks across story duration

Results

Average proportion of looks

Figure 2 shows the average proportion of looks to the target referent across *Time Bin,* where the zero line indicates the referring expression onset and each time bin represents ~83 ms of time. The proportion of looks were averaged over age group (children versus adults separately) and items (NPs versus pronouns separately). For the NPs (left panel) you can see that the proportion of looks to the target increased following referring expression onset and more so for the children (dashed line) compared to adults (solid line). For the pronouns (right panel) the relationship is less clear. However, it should be noted that there were many fewer pronouns than NPs, which can also be seen in the relative smoothness of the lines. Figure 3 shows the average proportion of looks to the target across the duration of the story (*Story Position*), averaged over age group and time bin. For both NPs (left panel) and pronouns (right panel), as well as children (triangles) and adults (circles) there is a clear downward linear trend, meaning that the overall proportion of looks to the target decreased as the story progressed over time.

Analysis

The gaze data were analyzed in R (version 3.1.2; R Core Team 2014) using Generalized Additive Mixed Modeling (GAMM, Wood 2006, mgcv R-package). GAMM is a nonlinear regression method that allows for the modeling of both linear and nonlinear random effects. We opted to use GAMMs over more standardized linear modeling because GAMMs are specifically designed to model nonlinear data and like most time series data, eye-tracking data is almost always nonlinear (Porretta, Kyröläinen, van Rij, & Järvikivi, 2018). The nonlinear relationship between the dependent variable and the predictors is modeled as a smooth function, which is a weighted sum of a set of base functions that each have a different shape. Using logistic GAMMs, we analyzed the counts (looks to the target vs. looks elsewhere) for each time bin (see Porretta et al. for discussion of binomial GAMMs for eye tracking data).

To determine the best-fitting model we did not perform a model comparison procedure, as model comparisons are not very reliable for binomial data (Wood, 2017). Instead, we included binary predictors (which model the difference between conditions) so that we could use summary statistics provided by the mgcv package to determine the significance of the smooth terms. In addition, we used visualization methods to interpret and verify the contribution of the smooth terms (cf van Rij, Hollebrandse, & Hendriks, 2016; van Rij, Hendriks, van Rijn, Baayen, & Wood, in press).

GAMM model of target looks

In order to investigate the eye gaze patterns between the four experimental conditions (adult NPs, child NPs, adult pronouns and child pronouns), we created three binary predictors. *IsChild*, which models the difference between adults (reference level) and children, *IsPronoun*, which models the difference between noun phrases (reference level) and pronouns, and *IsChildPronoun*, which models the additive interaction effect between *IsChild* and *IsPronoun*. We then included the predictor *Time Bin,* in order to analyze looks to the target referent across the time bin analysis window (where time bin 0 was the referring expression onset). We also included the predictor *Story Pos* (i.e., how far into the story the referring expression occurred), in order to test whether looks to the target referent changed as the story progressed. All predictors were allowed to interact. The smooth functions (s()) model the nonlinear regression lines for *Time Bin* and *Story Pos* interacting with the four experimental conditions. The nonlinear tensor product interactions (ti()) model the nonlinear interaction surface between *Time Bin* and *Story Pos* and the four experimental conditions, allowing us to

investigate whether the gaze patterns relative to hearing the referring expression change over the course of the story. To account for individual variation between participants, we included nonlinear random by-Subject factor smooths for *Time Bin* and *Story Pos*, as well as a random intercept for Event (unique Subject-Item combination). To account for autocorrelation in the residuals, an AR1 model was included by specifying the rho parameter and starting point for each time series (cf. Baayen, van Rij, de Cat, & Wood, 2018; van Rij et al., in press). The full analysis can be found in the supplementary materials and the final model summary is presented in Table 1.

Table 1: Summary of the partial effects in GAMM fitted to count data (looks to target vs. looks elsewhere)

A. parametric coefficients	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	-1.8000	0.2809	-6.409	1.47e-10 ***
B. smooth terms	EDF	Ref.df	Chi.sq	p-value
s(Time Bin)	4.156	4.626	7.369	0.12902
s(Time Bin):IsChild	4.436	4.834	3.067	0.68953
s(Time Bin):IsPronoun	6.134	7.426	19.762	0.00653 **
s(Time Bin):IsChildPronoun	8.908	9.618	85.969	4.87e-14 ***
s(Story Pos)	1.001	1.001	7.169	0.00743 **
s(Story Pos):IsChild	1.000	1.000	0.410	0.43476
s(Story Pos):IsPronoun	2.599	2.650	7.079	0.14016
s(Story Pos):IsChildPronoun	1.005	1.006	1.066	0.45868
ti(Time Bin, Story Pos)	14.429	15.616	130.820	< 2e-16 ***
ti(Time Bin, Story Pos):IsChild	14.153	15.493	126.287	< 2e-16 ***
ti(Time Bin, Story Pos):IsPronoun	14.677	15.613	182.723	< 2e-16 ***
ti(Time Bin, Story Pos):IsChildPronoun	14.448	15.444	153.732	< 2e-16 ***
s(Time Bin, Subject)	174.863	241.000	2946.556	< 2e-16 ***
s(Story Pos, Subject)	38.477	241.000	10609.166	0.00118 **
s(Event)	986.692	1218.000	6713.175	< 2e-16 ***

For the reference level (adult NPs) there was a significant nonlinear interaction between *Time Bin* and *Story Pos* (Chi.sq(14.429)=130.82; p<.001), meaning that target looks relative to hearing an NP changed as the story progressed over time. The interaction between *Time Bin* and *Story Pos* was also significant for each binary predictor: *IsChild* (Chi.sq(14.153)=126.29; p<.001), *IsPronoun* (Chi.sq(14.677)=182.72; p<.001) and *IsChildPronoun* (Chi.sq(14.448)=153.73; p<.001). Thus, we can conclude that all four experimental conditions have unique interaction surfaces. In order to interpret the interactions, we must use visualization. The contour plots in Figure 4 show how target looks relative to hearing an NP changed as the story progressed over time for both adults and children. The plots can be read like a topographic map with peaks and valleys, where pink indicates more looks to the target and green indicates more looks elsewhere. Both adults' and children's target looks increased after hearing an NP; however, this likelihood decreased in a nonlinear fashion as the story progressed over time. For example, approximately 30 seconds into the story (y-axis), it can be seen that the color changes from green to pink, moving from left to right (x-axis), indicating that target looks began increasing around time bin 3 (~250 ms after NP onset). However, 250 seconds into the story (y-axis), it can be seen that there is relatively solid green, moving from left to right (x-axis), indicating that there was almost no effect of Time Bin later on in the story. It can also be seen that the peak is steeper for children as compared to adults. The contour plots in Figure 5 show how target looks relative to hearing a pronoun changed as the story progressed over time for both adults and children. The white bands indicate the places throughout the story where there are no data, and thus should not be taken into consideration when interpreting the interaction surface. Similar to the NPs, both adults' and children's target looks increased after hearing a pronoun, but again the likelihood decreased in a nonlinear fashion as the story progressed over time. Reflected by the overall color becoming greener as you move from the bottom to the top of the plots (y-axis). It also appears that earlier on in the story (~30 seconds on the y-axis) the increase in target looks happens sooner for adults than it does for children; around time bins 3 (~250 ms) and 8 (~667 ms) respectively. However, because there were a lot less data for the pronouns we need to be careful to avoid over-interpretation.

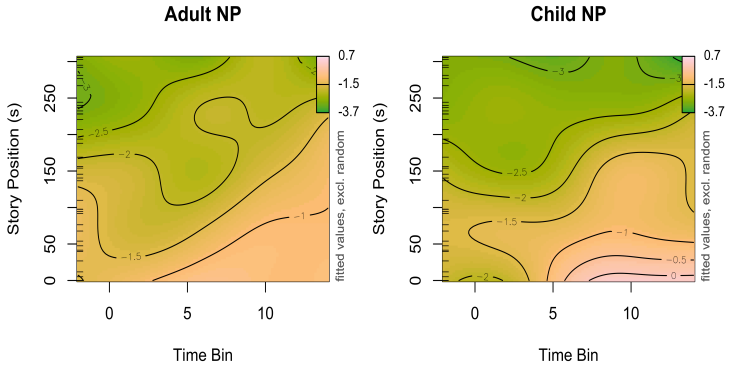


Figure 4: NP contour plots of the interaction between *Time Bin* and *Story Position* for adults and children. Pink indicates more target looks and dark green indicates fewer target looks.

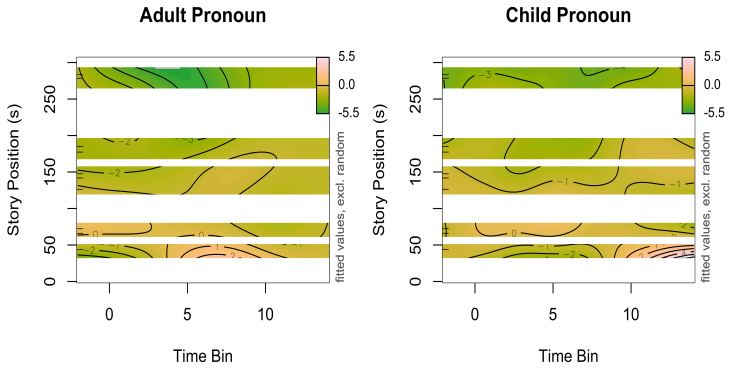


Figure 5: Pronoun contour plots of the interaction between *Time Bin* and *Story Position* for adults and children. Pink indicates more target looks and dark green indicates fewer target looks.

Discussion

The current study applied a novel adaptation of the visual world eye-tracking paradigm (VWP) in order to investigate the online comprehension of reference in a naturalistic language setting.

Overall, we found that eye gaze patterns relative to the onset of referring expressions (full NPs and pronouns) were largely influenced by when in the story the referring expressions occurred. When participants (both children and adults) heard a referring expression earlier on in the story, there was an increase in looks to the target referent. However, when participants heard a referring expression later on in the story there was no increase in looks to the target referent. This finding is in line with that of Engelen et al. (2014), who also found that overall target fixations following a referring expression decreased across a continuous discourse. They suggested that the visual scene may be particularly useful when first building a mental representation of the discourse, such that listeners search for appropriate referents in the visual scene, which results in eye movements being closely time-locked with the unfolding linguistic input. However, once a mental representation is well established, the visual scene does not provide any additional information and therefore eye movements are not closely time-locked. The argument for the visual scene not providing any additional information may be particularly relevant in the case of Engelen et al. (2014), given that the same visual scene was on display for their entire 7-minute discourse. As such, we originally proposed that their findings may be an artifact of fatigue or boredom. Interestingly, we found the same pattern despite using 22 different visual scenes throughout a 5-minute discourse. Based on the similarity of findings, we no longer believe that it is the type of visual scene (simple versus more complex) that causes the downward trend, but more likely a difference in the role that the visual scene plays throughout continuous discourse processing.

Although the visual context in the present study differed from that of Engelen and colleagues (2014), in both studies only a single mental representation of the linguistic discourse was required. This differs from more traditional VWP studies, in which items are presented in isolation and therefore participants must build a new mental representation for each item. In these studies, the eye gaze patterns associated with each item reflect the same type of processing that is, trying to understand who is doing what to whom (i.e., constructing a mental representation). But, because there is no additional linguistic context, the visual scene becomes particularly important for extracting information. This results in a close time-locking between linguistic input (i.e., the mention of a referring expression) and corresponding eye movements in the visual scene (i.e., looking at the referent almost immediately after it being mentioned). In our study, we also saw this for items that occurred earlier on in the discourse. So why did we not see it for items that occurred later on? Perhaps it is simply because later on in the discourse participants already know who the referents are and generally what is going on. In other words, participants already have an established mental representation of the discourse. Therefore, they do not rely on the visual scene for information to the same extent as they do at the beginning of the discourse (or in the more traditional VWP experiments that use thematically mutually unrelated 1-3 sentence stimuli). This results in eye movements not being closely time-locked with the linguistic input in the same way that they are at the beginning of the discourse and in more traditional VWP studies. It is not that eye gaze patterns later on in the discourse are arbitrary (or unmeaningful), but rather that they may reflect a type of processing for which the timing between the linguistic input and the corresponding eye movements and gaze location is not yet well understood. One possibility is that, as the discourse status of a referent changes due to repeated mentions and due to the continuous story context, participants engage in inspecting other aspects of the visual scene to refine their discourse representation, instead of looking at the referent each time it is mentioned.

In addition to looking at overall eye gaze patterns, we also compared the eye gaze patterns between full NPs and pronouns, as well as between children and adults. Following the mention of an NP, we found that earlier on in the story there was a greater increase in the proportion of looks to the target referent for children compared to adults. However, as the story unfolded children became less likely to fixate on the target compared to adults (i.e., there was a stronger effect of story position for children). Following the mention of a pronoun, we found that earlier on in the story children and adults showed a similar increase in the proportion of looks to the target referent, but this happened sooner for adults than children (~250 versus ~667 ms after pronoun onset, respectively). However, given that the dataset was relatively limited, these findings are preliminary and invite further research.

Given the novelty of the present study there were several challenges, the primary one being the technical issues with the eye-tracking glasses, which resulted in >50% of the children being excluded from the analysis. Furthermore, we ended up having a lot fewer referring expressions to analyze than we would have liked (especially in the case of pronouns). This was because our primary goal was to keep the story as natural-sounding as possible and including an excessive amount of referring expressions would have been counterproductive to this goal. Together, these two factors resulted in there being a relatively small dataset, which always runs the risk of lacking statistical power. Nonetheless, the findings from the current study build upon those reported by Engelen et al. (2014) and provide convincing evidence that the relationship between linguistic input and gaze behavior is heavily influenced by context. They further suggest that this relationship is affected by the discourse status of the referent, which changes over the course of a normal continuous story. Furthermore, the findings demonstrate the importance of investigating language processing under naturalistic conditions. Future research is needed to better understand the link between linguistic input and corresponding eye movements.

Acknowledgments

We would like to acknowledge the Social Sciences and Humanities Research Council of Canada (SSHRC), as well Words in the World (a SSHRC partnered research training initiative) for funding this research. We would also like to thank Kaleigh and Romy for all their help with the creation and recording of the experimental stimuli.

References

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The immediate use of gender information: Eye-tracking evidence of the time-course of pronoun resolution. Cognition, 76, B13–B26.

Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. (2018). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), Mixed-Effects Regression Models in Linguistics (pp. 49-69). (Quantitative Methods in the Humanities and Social Sciences). Springer International Publishing AG.

Cooper, R. M. (1974). The Control of Eye Fixation by the Meaning of Spoken Language: A New Methodology for the Real-Time Investigation of Speech Perception, Memory, and Language Processing. Cognitive Psychology, 684-107.

Crawley, R., Stevenson, R., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. Journal of Psycholinguistic Research, 4, 245–264.

Engelen, J. A., Bouwmeester, S., de Bruin, A. B., & Zwaan, R. A. (2014). Eye movements reveal differences in children's referential processing during narrative comprehension. Journal Of Experimental Child Psychology, 11857-77.

Hartshorne, J. K., Nappa, R., & Snedeker, J. (2015). Development of the first-mention bias. Journal of child language, 42(2), 423-446.

Järvikivi, J., Pyykkönen-Klauck, P., Schimke, S., Colonna, S., & Hemforth, B. (2014). Information structure cues for 4-year-olds and adults: Tracking eye movements to visually presented anaphoric referents. Language, Cognition And Neuroscience, 29(7), 877-892.

Järvikivi, J., van Gompel, R. P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. Psychological Science, 16(4), 260-264.

Kaiser, E., & Trueswell, J. C. (2008). Interpreting Pronouns and Demonstratives in Finnish: Evidence for a Form-Specific Approach to Reference Resolution. Language And Cognitive Processes, 23(5), 709-748.

Noldus Information Technology. (2012). The Observer XT reference manual 11.0. Wageningen, the Netherlands: Author.

Porretta, V., Kyröläinen, A., van Rij, J., & Järvikivi, J. (2018). Visual world paradigm data: From preprocessing to nonlinear time-course analysis. In Czarnowski I, Howlett R and Jain L (eds.), Intelligent Decision Technologies 2017, number 73 series Smart Innovation, Systems and Technologies, pp. 268–277.

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

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

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

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. Science, (5217), 1632.

van Rij, J., Hollebrandse, B., Hendriks, P., (2016). Children's Eye Gaze Reveals their Use of Discourse Context in Object Pronoun Resolution. In: Holler A. Glauw C. Suckow K. (eds.) *Empirical Perspectives on Anaphora Resolution*. Berlin: Mouton de Gruyter.

van Rij, J., Wieling, M., Baayen, R.H., & van Rijn, H. (2015). itsadug: interpreting time series and autocorrelated data using GAMMs.

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R.H., & Wood, S.N. (Accepted for publication in Trends in Hearing Science). Analyzing the time course of pupillometric data.

Wood, S. (2017). Generalized Additive Models. New York: Chapman and Hall/CRC.

Wood, S. N. (2006). Generalized additive models: an introduction with R (Vol. 66). Boca Raton: Chapman & Hall/CRC Press.