Investigating The Effects Of Visual And Linguistic Context On Object Pronoun Processing As Measured By Pupil Dilation And EEG

Bachelor's Project Thesis

Tineke Jelsma, s2719576, t.jelsma@student.rug.nl,
Supervisors: dr. J.P. Borst & dr. J.C. van Rij-Tange

Abstract: This study aimed to investigate whether visual and linguistic context affect the processing of object pronouns ('him', 'her'). Recent studies have shown, by measuring pupil dilation, that context information affects object-pronoun processing in an early stage and interacts with grammatical processing (e.g., van Rij et al, 2016; van Rij, 2012). In the current study, we investigated the visual and linguistic effects by measuring the exact timing of those contexts on object-pronoun processing by co-registering EEG and pupil dilation. With a 2x2x2 within-subject design, we investigated the effects of visual context (other-oriented vs self-oriented action; e.g., a picture with a hedgehog photographing a mouse, or a hedgehog photographing himself), discourse prominence (the introduction sentence introduces the actor first or second; 'You just saw a hedgehog and a mouse.' vs 'You just saw a mouse and a hedgehog.'), and referring expression (the test sentence contains an object pronoun or a reflexive (fillers); 'The hedgehog was photographing him / himself with a camera.'). Surprisingly, we found that visual and linguistic context do not have an influence on object-pronoun processing. However, we did find an influence of visual context on the processing of reflexives in EEG and an interaction between visual and linguistic context in pupil dilation.

1 Introduction

Pronouns like *him* and *himself* are used in language to refer to a referent. Pronouns do not have a specific interpretation. In Dutch, only the gender and number directly limit the choice of referents (*him* would refer to a male singular referent). Often multiple characters are used in a story. Therefore, listeners have to determine the character the pronoun is referring to. Listeners use preceding language to determine their choice. This can be the grammatical role (same number and gender), order of mention (earlier mentioned objects are preferred over later mentioned objects), and recency (the object should be mentioned recently) (see Arnold, 1998, for an overview).

Object pronouns are pronouns that are used as objects of a sentence, for example: "The bear tickled *him* with a feather." In both Dutch and English it is impossible that the object pronoun refers to the subject (the bear). This restriction is known as Principle B of Binding Theory (Chomsky, 1981).

The initial-filter account (a.o., Clifton, Kennison, & Albrecht, 1997; Nicol & Swinney, 1989; Chow et al., 2014) assumes that listeners initially only apply grammatical principles to determine the correct referent and that the linguistic discourse only plays a role after excluding potential referents that violate Principle B.

However, several studies suggest that not only grammar, but also the linguistic discourse initially influence adults' online interpretation of object pronouns (Spenader et al., 2009; Clackson, Felser, & Clahsen, 2011). This is support for the competing-constraints account (a.o., Badecker & Straub, 2002; Kennison, 2003), which assumes that grammar competes with other sources of information (linguistic discourse) during interpretation of the object pronoun.

One of the recent studies supporting this (van Rij, 2012) reports a combined effect of visual context and linguistic context on object pronoun processing. Participants saw an image of an actor and a patient. Then they listened to the introduction of the referents, either actor first or second, (linguistic context) and the image was described. The description could be either congruent or incongruent with the image (visual context). Effects were found on pupil dilation 500-1000 ms after the object pronoun onset. The pupil dilated, which indicates that cognitive processing load is higher (a.o., Hyönä et al., 1995). The results suggest that when the agent is introduced first, the visual context affects the pronoun processing. When the agent is introduced second, the visual context does not affect the pronoun processing. This study provides evidence against the initial-filter account, as not only grammar, but also linguistic discourse and visual context influence object pronoun processing.

However, pupil dilation is a relatively slow measurement. Effects of a pupil dilation are found on average around 1000 ms after the stimulus that triggered the dilation (e.g., Hoeks & Levelt, 1993). A more precise measurement tool would give more information on the timing of the visual and linguistic context. This allows to distinguish between the two processing accounts. Therefore, EEG would be more appropriate.

In this study, we investigated the effects of visual and linguistic context on object pronoun processing. Our hypothesis is that the discourse (visual and linguistic context) sets up an expectation for an interpretation that may be congruent with the pronoun interpretation or not. The introduction order modulates the expectation: When the actor is introduced first, the expectation is stronger than when the actor is introduced second. This is caused by the order of mention effect: it feels natural to have the actor introduced first, so with the actor introduced second, the listener will already be surprised and is less surprised if the story is incongruent as well. If the story is incongruent with the image, we expect to see an increase in pupil dilation, this represents more processing load. For the EEG we expect to see an N400 effect, which is associated with meaningful, semantical stimuli.

We operationalized this by using both pupil dilation and EEG. The pupil dilation allows us to compare the results with van Rij's results. The EEG gives us a better temporal resolution. A difference between our study and van Rij's study is that van Rij used a picture verification task without a blank screen and we added a blank screen. The blank screen paradigm (Altmann, 2004) removes the image off the screen before the spoken description starts. Altmann showed that the physical presence of picture is not necessary; listeners use a mental presentation of the image. By using the blank screen paradigm, we try to minimize the involvement of the visual system in language processing. This results in a cleaner EEG.

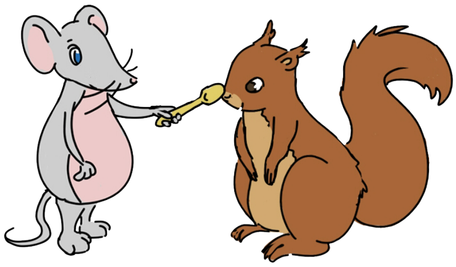

Figure 2.1: Example of an other-oriented picture.

2 Methods

2.1 Participants

In total, 32 students (right-handed, native Dutch speakers) participated in our experiment of whom 18 were male and 14 female. The average age of the participants was 21.5 years old, ranging between the ages of 19-24. All participants signed an informed consent form before the start of the experiment. On average, the total time spent on both setting up and conducting the experiment took approximately 1.5 hours. Participants received a compensation of €12,-.

2.2 Design

By combining a picture verification task with a blank screen paradigm, we investigated the influences of both visual and linguistic context on ob-

Figure 2.2: Example of a self-oriented picture.

ject and reflexive pronoun processing.
The 2x2x2 (Picture Type x Introduction Order x Referring Expression) design was tested within subjects and partly within items (four variants of each item were tested in each experimental session).
For Picture Type there were two conditions: other-oriented pictures (see Figure 2.1), in which the actor (an animal) is performing an action upon another animal, and self-oriented pictures (see Figure 2.2), in which the actor is performing the action upon himself. There were 80 unique pictures, which formed 40 pairs of two variants of the same picture. Introduction Order is related to the introduction sentence which followed the picture. There were two conditions: an actor-first introduction, in which the actor is mentioned first (*"Zojuist zag je een muis en een eekhoorn."*, you just saw a mouse and a squirrel) or actor-second (*"Zojuist zag je een eekhoorn en een muis."*, you just saw a squirrel and a mouse). After the introduction sentence, the test sentence was played, containing the referring expression. This could be either an object pronoun, *"hem"*, (*"De muis raakte hem aan met een lepel."*, the mouse touched him with a spoon) or a reflexive pronoun, *"zichzelf"* (*"De muis raakte zichzelf aan met een lepel."*, the mouse touched himself with a spoon). The reflexive sentences were included as fillers.
Each picture was shown twice to the participants, with each image occurring in a different block and with two of the four conditions (Introduction Order x Referring Expression). We tested multiple variations of each item to increase the amount of data necessary for EEG analysis. To avoid any repetitions of these combinations, or to avoid two of the same combinations after each other, we made unique lists for all participants. These lists consisted of four different blocks. For each participant, the order of items within each block was randomized to avoid any further bias.

2.3 Material/Stimuli

Stimulus presentation was programmed in Experiment Builder (SR Research, 2017).
The pictures consisted of a combination of the visual stimuli of van Rij (2012) and van Rij et al (2016). They were presented centrally against a light grey background with a width of 500 pixels. The height depended on the image ratio. 50% of the pictures were randomly selected and mirrored. For each picture, we recorded a two-sentence description. These sentences were recorded in the recording studio of the Faculty of Arts, University of Groningen, and afterwards manipulated, by means of splicing and normalizing, with the program PRAAT (Boersma & Weenink, 2018).
Two kinds of introduction sentences were recorded: actor-first and actor-second. They were all built in a similar style with artificial breaks: "Zojuist zag je" (you just saw) + 100 ms silence + <referent>+ 100 ms silence + "en" (and) + <referent>.
Three kinds of test sentences were recorded: with either an object noun sentence, an object pronoun or a reflexive pronoun. The sentence with an object noun sentence (e.g., *"De muis raakte de eekhoorn aan met een lepel."*, the mouse touched the squirrel with a spoon) was used as carrier phrase for the test sentences. The pronouns (*"hem"*) and reflexives (*"zichzelf"*) were spliced into these object noun parts, so the intonation of the rest of the sentence would be kept identical. The test sentence, too, had artificial breaks: <Actor>+ 100 ms silence + <verb>+ 100 ms silence + <pronoun/reflexive>+ 100 ms silence + <prepositional phrase>.
Between the introduction sentence and the test sentence was a fixed break of 200 ms.
The answer screen contained two boxes: one green, with the word "correct" in it, and one red, with the word "incorrect" in it. The order of these boxes was randomly determined for all trials in the experiment to prevent motor preparation. The Ctrl-left

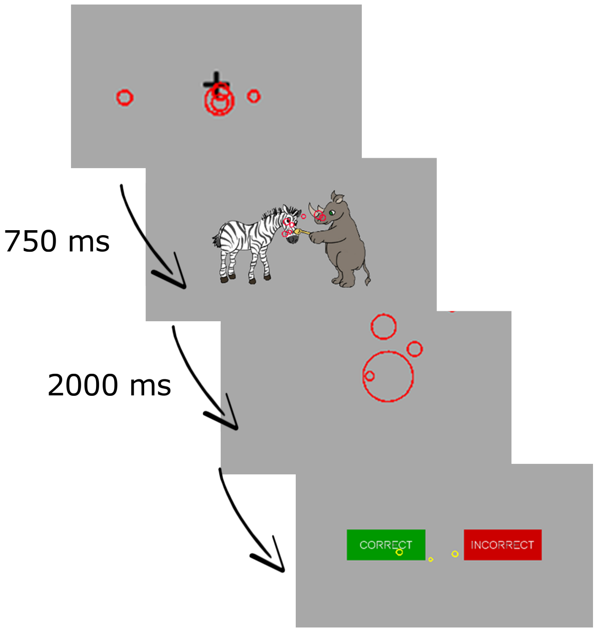

Figure 2.3: Visualization of a trial.

button was linked to the left-positioned answer, the Ctrl-right button to the right.
The pupil of the left eye was monitored continuously during the picture verification task with the EyeLink 1000 (SR research) at 500 Hz (16 mm lens + target sticker). Brain activity was measured via EEG caps consisting of 32 electrodes and six external electrodes: on the mastoids, HEOG and VEOG (above and below the right eye). These were connected to BioSemi, which recorded the data at 2048 Hz.

2.4 Procedure

In advance, the participant was informed to wear neither glasses nor mascara, since both influence the precision of the eye-tracker's pupil detection - hence, the eye-tracking results. Before going on to the actual experiment, the participant had to sign the consent form. The participant was positioned on a non-adjustable chair behind a computer screen, which was positioned on a desk. On the ground, the target position for the chair was indicated using tape, to make sure all participants would be facing the eye-tracking from a fixed location. The eye-tracker was installed in front of the computer, at a distance of approximately 70 centimeters to the participant's eyes. The keyboard was positioned between the eye-tracker and the participant, at a distance comfortable for the participant.
After this part of the set-up, the EEG cap and electrodes were positioned. On the participant's forehead, we placed a target sticker (i.e., a sticker with a bullet point on it), for the eye-tracker to detect as target point. An oral instruction was followed by a similar instruction on the screen: during this instruction, participants became familiar with the kind of pictures they were about to face and the keys they had to press accordingly. Then, they had to perform an eye-tracker calibration followed by a validation. We aimed for an average deviation value of 0.5. If it was higher than 0.5, another calibration had to be conducted (the lower this value, the more precise the calibration and thus the eye-tracking data).
Figure 2.3 visualizes the structure of the trial. Each trial started with a fixation point. An invisible square surrounded this fixation point and the participant had to look for at least 100 ms within this square to start with the trial. 650 ms after this, the picture appeared on the screen, and was shown for 2000 ms. A blank screen followed. 500 ms later, while still seeing the blank screen, the two-sentence stories were played. 1200 ms after the offset of the test sentence, the answer screen appeared during which the participant had to indicate whether the story was congruent with the picture or not. They had 5 seconds to give an answer.
If the eye tracker did not recognize that the participant was looking within the invisible square surrounding the fixation point for 100 ms within 5 seconds, another calibration had to be performed and the trial would be skipped.
The participant started with three practice trials with pictures that were not used during the actual experiment.
After the participant had finished the practice trials and asked all their possible questions, they would perform another calibration before continuing to the actual experiment. In total, the participant performed 160 trials, divided into 4 blocks of 40 trials. Each block was separated from the next one by a break. After each break, the participant was asked to calibrate again.
Luminance of the room was normal and kept constant during the experiment.

2.5 Analysis methods

The EEG data have been pre-processed with a script from Jelmer Borst, using EEGLAB (Delorme & Makeig, 2004). The data has been re-referenced to the mastoids electrodes and downsampled to 100 Hz. The low pass filter was set to 40 Hz, which removes fast noise. The high pass filter was set to 0.01 Hz, to remove very slow noise. After the downsampling and filtering, trials with extreme values were manually rejected. Blinks and saccades have been removed with ICA.

The pupil dilation data have been automatically pre-processed with a script from Jacolien van Rij, using R (R Core Team, 2018). Blinks and saccades have been removed from the data, with 100 ms padding around the blink, and 10 ms around the saccade.

The first 80 trials (that is, the first two blocks) have been used for the analysis, since then the participants had only encountered each picture once and these results were more reliable as the participants reported they became more distracted during the last two blocks.

The data was baselined on 250-0 ms before the pronoun onset in order to investigate differences between conditions starting from the pronoun onset. Statistical analyses have been performed on the eye-tracking and EEG data, using linear mixed-effects (LME) models. Even though the reflexive sentences were only meant as fillers, we have performed analyses on both sentences types.

For eye tracking, the window on which the analyses have been performed is 750 to 1250 ms after the onset of the pronoun. The reason for this is that pupil dilation peaks around 1000 ms after the stimulus onset that triggered the dilation. For each subject, the median pupil size per trial within this window has been taken and analyzed.

In EEG, we expected to see a N400 or a P600 when the trial is incongruent. The N400 is expected because it is related to semantic violations. However, if incongruency is only detected after grammar processing, an N600 is expected (syntactical violation). Therefore, we performed an analysis on two time windows: the first from 300 to 500 ms after the pronoun onset and another from 500 to 700 ms after the pronoun onset. Similar to the pupil dilation, the median Cz value has been taken for all the subjects per trial, followed by a comparison of the means of those.

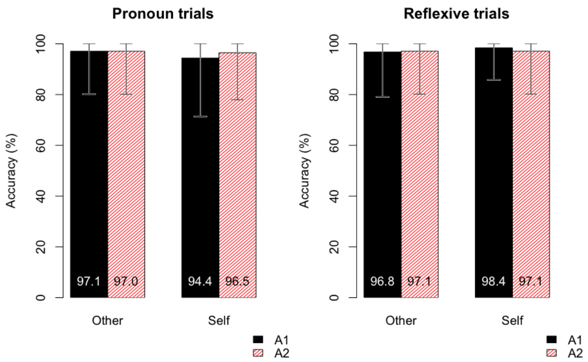

Figure 3.1: Accuracies and standard error of the trials, divided in other/self-oriented pictures and A1 & A2 introduction orders.

For all of the data, three models have been assessed: the simplest one only checking for main effects, the second also including all two-way interactions and the most complex one adds the three-way interaction to that. These two latter, more complex, models have only been used if they proved to explain significantly more variation than the simplest model. If the most complex model, which includes the three-way interaction, turned out to be the best model, two separate analyses have been performed on the two sentence types (pronoun and reflexive), apart from each other.

3 Results

In Figure 3.1 the mean accuracy and the standard deviation for all conditions are presented. Overall, the accuracy is very high, this was expected because the experiment was fairly easy. It shows that the participants paid attention to the trials.

3.1 Eye pupil dilation

In Figure 3.2 and 3.3 the data is visualized. The baseline is set 250 ms before the onset of the pronoun/reflexive. In the legends of the figures, the O stands for other-oriented action, the S for self-oriented action. A1 means that the actor is introduced first and A2 means that the actor is introduced second.

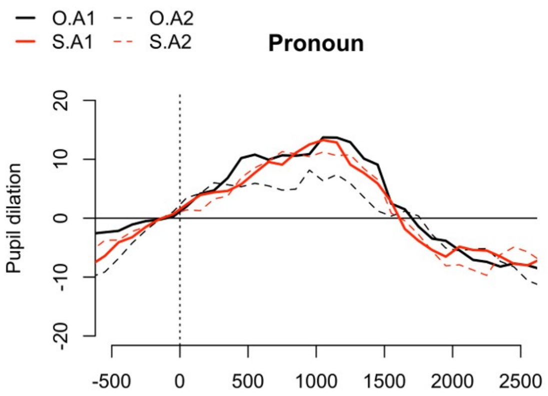

Figure 3.2: The pupil dilation of the pronoun trials.

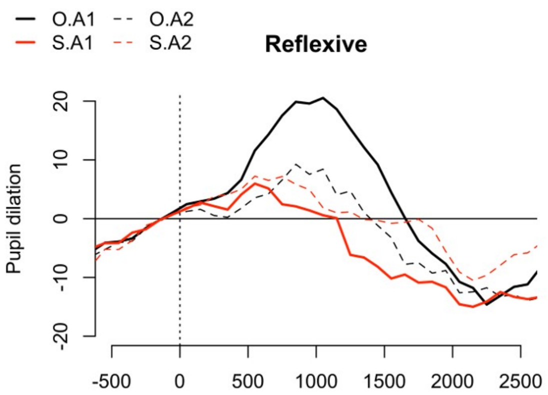

Figure 3.3: The pupil dilation of the reflexive trials.

In Figure 3.2 all the different conditions show an increase in pupil dilation after the onset of the pronoun, no specific effects or differences can be seen. In Figure 3.3 trials with an other-oriented action (incongruent with the reflexive) and with the actor introduced first (O.A1) have a big increase in pupil dilation in comparison with the other conditions. While the grand average of condition O.A1 reaches 20 in its peak, the grand averages of other conditions peak around 10.

3.2 Eye pupil dilation analysis

The variables used in the LME are the Introduction type, Picture type, and Sentence type. The best fitting model is found by using the backward fitting process. The three-way interaction effects model ($\chi^2(1)$ = 5.74, p = 0.017) is the significantly improved model. This model (see Table 1 in the appendix) shows significance for the effect of sentence type (for $\beta_{sentencetypeR}$ = 17.684, SE = 7.146, t = 2.475), the picture type:sentence type interaction ($\beta_{pictypeS:sentencetypeR}$ = -41.064, SE = 10.057, t = -4.083) and the introduction type:picture type:sentence type (three-way) interaction ($\beta_{introtypeA2:pictypeS:sentencetypeR}$ = 34.175, SE = 14.255, t = 2.397).
The analysis is further split into pronoun models and reflexive models, as some effects might only occur in one of them. None of the models for the pronoun data show significance. The two-way interaction model for the reflexive data is the best fitting model ($\chi^2(1)$ = 11.097, p <0.01). In this model, significance is found in the main effect of introduction type ($\beta_{introtypeA2}$= -23.140, SE = 6.957, t = -3.326), main effect of picture type ($\beta_{pictypeS}$ = -32.672, SE = 6.865, t = -4.760) and an interaction between those effects ($\beta_{introtypeA2:pictypeS}$ = 32.603, SE = 6.865, t = 3.340). These effects tell that there is a stronger effect of introduction order on other-oriented pictures than on self-oriented pictures.

3.3 EEG

In Figures 3.4 and 3.5 the results of the EEG data are presented. The data is from the Cz electrode, which is positioned on the middle line of the head and central. The legend for the EEG results is the same the legend for the pupil dilation. The

electrical activity is measured in micro volt.
In Figure 3.4 the results of the pronoun trials in EEG are presented. Similar to the eye pupil dilation results of pronoun trials, no difference between the conditions is visible.
The graph in Figure 3.5 shows a difference between the conditions. Both the trials with an other-oriented picture (incongruent with the reflexive), regardless of the introduction order, show a higher peak at around 600 ms after the onset of the reflexive.

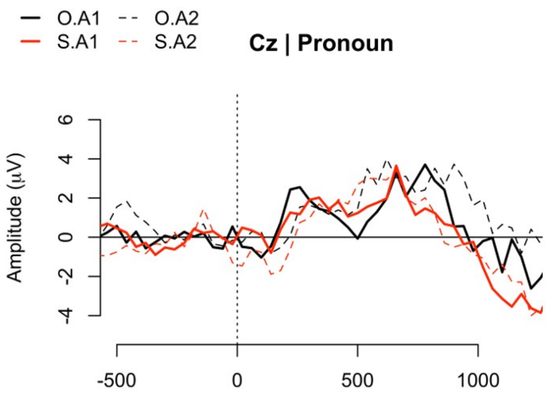

Figure 3.4: EEG data of the pronoun trials.

3.4 EEG analysis

For the EEG analysis, we also applied the Linear Mixed Effect model on both object pronouns and reflexives. The analysis for the N400 has been applied to 300-500 ms after the onset of the pronoun/reflexive. With the backward fitting process, no significant interactions (two-way or three-way) are found. In the main effects model (see Table 3 in the appendix), no significant fixed effects are present.
We also analyzed the time frame of 500-700 ms after the onset of the pronoun/reflexive, because a peak around 600 ms after the onset of the pronoun can be seen in Figure 3.5. With the backward fitting process, the best fitting model is the main effects model (see Table 4 in the appendix). However, there is one interaction in the two-way interaction model (see Table 5 in the appendix) a significant interaction of the picture type:sentence type ($\beta_{pictypeS:sentencetypeR}$ = -3.31080, SE = 1.47590, t = -2.243) can be found. This effect has been combined with the main effects model. This new model is compared with the main effects model and the new model is the best fitting model (($\chi^2(1) = 4.97$, p = 0.026). This model (see Table 6 in the appendix) shows significance in some fixed effects: The sentence type ($\beta_{sentencetypeR}$ = 2.68, SE = 1.042, t = 2.571) and the picture type: sentence type interaction ($\beta_{pictypeS:sentencetypeR}$ = -3.30, SE = 1.476, t = -2.36). Thus, only for the reflexive sentences, we have found an effect of picture type.

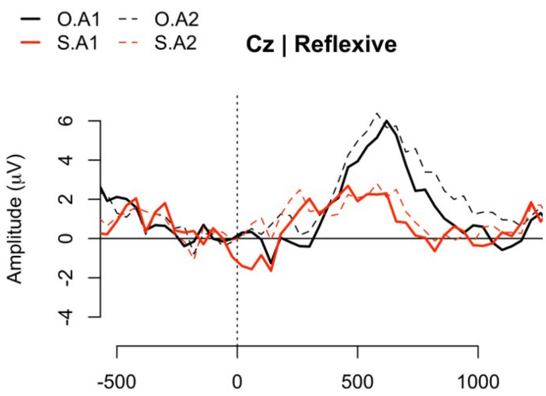

Figure 3.5: EEG data of the reflexive trials.

3.5 Summary of the results

Overall, no significant effects have been found in the pronoun data. This is in contrast with our hypothesis. In the results of the reflexive trials, different significant effects have been found. For the eye pupil dilation, an effect of introduction type (A2), an effect of picture type (self-oriented picture) and an interaction effect between those effects are significant in the reflexive trials.
For the EEG, no N400 effects were found (around 300-500 ms after the onset of the reflexive), but around 500-700 ms after the reflexive (P600 effect), the picture type effect was significant.

4 Discussion

This study investigated the influence of linguistic and visual context on object pronoun processing with a picture verification task. We hypothesized that with an incongruent picture and/or the actor introduced second eye pupil dilation would increase and EEG would show an N400 effect.
The results show that for pronouns there are no significant effects. For reflexive trials an interaction between picture type and introduction order was found in the eye pupil data. The EEG data show an effect of picture type for the reflexive trials. Therefore, results do not fit with the hypothesis.
While we did expect to see effects in the pronoun trials, there were none. Next to the principle B theory, one possible explanation is that the pronoun allows several referents while the reflexive allows one (the subject of the sentence). Participants were not set on one option and therefore the surprise effect is less in the pronoun trials compared to the reflexive trials. Van Rij's research (2012) did find effects of linguistic and visual context on object pronoun processing. A major difference between the two studies is the use of a blank screen in the trial. The blanks screen created extra time between the picture and the onset of the pronoun. It might be interesting to analyze the time before the onset of the pronoun. One more difference was the artificial pauses between the word groups, which has created a more artificial sound.
Another unexpected and surprising effect is the difference of the results in EEG and pupil dilation in the reflexive trials. While the pupil dilation analysis shows an interaction between the picture type and the introduction order, the EEG analysis only shows an effect of the picture type. EEG and pupil size are different measures, which are sensitive to different factors (e.g. blinks, saccades and muscle activity). This might have caused the difference. Apart from the measurement differences, a logical explanation of the eye pupil dilation data could be that when the actor is introduced second, the participant is already cautious and is not as surprised when the picture and description are incongruent as when the introduction order is presented with the actor first. The peak of condition O.A does not seem to support the initial filter account (a.o., Clifton, Kennison, & Albrecht, 1997; Nicol & Swinney, 1989; Chow et al., 2014).
In the EEG results, we found a picture type effect in the time frame around 600 ms after the onset of the reflexive. This indicates a P600 effect, while we expected to see a N400. An N400 is related to semantical errors, which would prove that not only grammar, but also other effects (visual and linguistic context) influence pronouns. P600 is relevant for language processing as well, as it is related to grammatical errors and syntactic errors. This might indicate that the difference between a pronoun and a reflexive is seen as a grammatical error, instead of a semantical error.
Some problems were encountered while running the experiment. Participants found this experiment to be a dull experiment. Some participants almost fell asleep towards the end of the experiment. There were also some problems with the pictures: when the animals had the same shape or color, participants found it harder to distinguish them, some animals and objects were ambiguous. The other-oriented pictures were easier to look at, because the action looked bigger in the picture.
We also assumed that the actor was the animal that was performing the action. However, some pictures described an action where an animal was pointing to another animal. It can be argued that the animal that was pointed to (the patient) looks more prominent in such a picture. This might have influenced the effect of linguistic context.
Our results give new opportunities for further research. It might be interesting to analyze the collected data in other time frames as well (before the onset of the pronoun/reflexive).
As our results were difficult to explain, it is

interesting to set up a similar experiment but with different trials, as several problems arose from the pictures. This might clarify the results in our research.

4.1 Conclusion

Using pupil dilation and EEG, we did not find effects on object pronoun processing. We did find effects of visual context on object reflexive processing in EEG (more processing load for incongruent items) around P600, and an interaction effect of picture type and introduction type in eye pupil dilation (incongruent trials with A1 introduction show more processing load). Overall, we did not find evidence against the initial filter account for object pronoun processing. In reflexive processing we found a P600 effect for picture type, which suggested that the incongruent pictures elicited an ungrammatical interpretation.

5 Acknowledgements

I would like to thank Anouk Hoekstra for recording the sentences, Robbert Prins and Petra van Berkum for drawing the pictures, and Jacolien van Rij and Jelmer Borst for supervising this project and contributing to the preprocessing and analyses.

References

Gerry TM Altmann. Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, 93(2):B79–B87, 2004.

Jennifer E Arnold. *Reference form and discourse patterns*. PhD thesis, Stanford University Stanford, CA, 1998.

William Badecker and Kathleen Straub. The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4):748, 2002.

Paul Boersma and David Weenink. Praat: doing phonetics by computer [computer program]. version 5.4.22., 2018. URL `http://www.praat.org/`.

Noam Chomsky. *Some concepts and consequences of the theory of government and binding*, volume 6. MIT press, 1982.

Wing-Yee Chow, Shevaun Lewis, and Colin Phillips. Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*, 5:630, 2014.

Kaili Clackson, Claudia Felser, and Harald Clahsen. Children's processing of reflexives and pronouns in english: Evidence from eye-movements during listening. *Journal of Memory and Language*, 65(2):128–144, 2011.

Albrecht J Clifton C, Kennison S. *Reading the words her, his, him: Implication for parsing principles based on frequency and on structure*, volume 36(2). Journal of Memory and Language, 1997.

A. Delorme and S. Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics. *Journal of Neuroscience Methods*, 134:9–21, 2004.

Bert Hoeks and Willem JM Levelt. Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*, 25(1):16–26, 1993.

Jukka Hyönä, Jorma Tommola, and Anna-Mari Alaja. Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology Section A*, 48(3):598–612, 1995.

Shelia M Kennison. Comprehending the pronouns her, him, and his: Implications for theories of referential processing. *Journal of Memory and Language*, 49(3):335–352, 2003.

Janet Nicol and David Swinney. The role of structure in coreference assignment during sentence comprehension. *Journal of psycholinguistic research*, 18(1):5–19, 1989.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL `https://www.R-project.org/`.

Jennifer Spenader, Erik-Jan Smits, and Petra Hendriks. Coherent discourse solves the pronoun interpretation problem. *Journal of child language*, 36(1):23–52, 2009.

SR Research Experiment Builder 2.1.140 [Computer software]. SR Research Ltd, Mississauga, Ontario, Canada, 2017.

Jacolien van Rij. *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults*. Rijksuniversiteit Groningen, 2012.

A Appendix

Table 1: Formula: medianPupil ~ (introtype + pictype + sentencetype)ˆ3 + (1 |Subject) + (1 |Item)

```
Random effects:
  Groups   Name        Variance Std.Dev.
  Item     (Intercept)   50.43   7.102
  Subject  (Intercept)  347.13  18.632
  Residual             7245.97  85.123
 Number of obs: 2289, groups:  Item, 40; Subject, 31

 Fixed effects:
                                     Estimate Std. Error t value
 (Intercept)                           17.905      6.175   2.900
 introtypeA2                           −5.064      7.127  −0.711
 pictypeS                               8.478      7.140   1.187
 sentencetypeR                         17.684      7.146   2.475
 introtypeA2:pictypeS                  −1.458     10.095  −0.144
 introtypeA2:sentencetypeR            −18.022     10.116  −1.782
 pictypeS:sentencetypeR               −41.064     10.057  −4.083
 introtypeA2:pictypeS:sentencetypeR    34.175     14.255   2.397
```

Table 2: Formula: medianPupil ~ (introtype + pictype)ˆ2 + (1 |Subject) + (1 |Item)

```
Random effects:
  Groups   Name        Variance Std.Dev.
  Item     (Intercept)   45.42   6.739
  Subject  (Intercept)  174.88  13.224
  Residual             6800.14  82.463
 Number of obs: 1147, groups:  Item, 40; Subject, 31

 Fixed effects:
                      Estimate Std. Error t value
 (Intercept)            35.663      5.536   6.442
 introtypeA2           −23.140      6.957  −3.326
 pictypeS              −32.672      6.865  −4.760
 introtypeA2:pictypeS   32.603      9.761   3.340
```

Table 3: (N400) Formula: Cz ∼ (introtype + pictype + sentencetype) + (1 |Subject) + (1 |Item)

```
Random effects:
  Groups   Name        Variance  Std.Dev.
  Item     (Intercept)   0.7624   0.8732
  Subject  (Intercept)   4.1170   2.0291
  Residual             143.9710  11.9988
 Number of obs: 1360, groups:  Item, 40; Subject, 21

 Fixed effects:
               Estimate Std. Error t value
 (Intercept)    1.57271    0.80901   1.944
 introtypeA2   -0.08103    0.65172  -0.124
 pictypeS      -0.04663    0.65188  -0.072
 sentencetypeR  0.56103    0.65518   0.856
```

Table 4:(P600) Formula: Cz ∼ (introtype + pictype + sentencetype) + (1 |Subject) + (1 |Item)

```
Random effects:
  Groups   Name        Variance Std.Dev.
  Item     (Intercept)   3.767   1.941
  Subject  (Intercept)   3.681   1.919
  Residual             181.047  13.455
 Number of obs: 1360, groups:  Item, 40; Subject, 21

 Fixed effects:
               Estimate Std. Error t value
 (Intercept)     3.1454     0.9080   3.464
 introtypeA2     0.8748     0.7313   1.196
 pictypeS       -1.7270     0.7316  -2.361
 sentencetypeR   1.0356     0.7418   1.396
```

Table 5: (P600) Formula: Cz ∼ (introtype + pictype + sentencetype)ˆ2 + (1 |Subject) + (1 |Item)

```
Random effects:
  Groups    Name         Variance  Std.Dev.
  Item      (Intercept)    3.479    1.865
  Subject   (Intercept)    3.718    1.928
  Residual               180.494   13.435
 Number of obs: 1360, groups:  Item, 40; Subject, 21

 Fixed effects:
                            Estimate  Std. Error  t value
 (Intercept)                 2.04204    1.12164    1.821
 introtypeA2                 1.37267    1.27535    1.076
 pictypeS                    0.08508    1.28682    0.066
 sentencetypeR               3.04516    1.27900    2.381
 introtypeA2:pictypeS       −0.21279    1.46014   −0.146
 introtypeA2:sentencetypeR  −0.71464    1.46041   −0.489
 pictypeS:sentencetypeR     −3.31080    1.47590   −2.243
```

Table 6: (P600) Formula: Cz ∼ introtype + pictype + sentencetype + pictype:sentencetype + (1 |Subject) + (1 |Item)

```
Random effects:
 Groups    Name         Variance  Std.Dev.
  Item      (Intercept)    3.481    1.866
  Subject   (Intercept)    3.701    1.924
  Residual               180.535   13.436
 Number of obs: 1360, groups:  Item, 40; Subject, 21

 Fixed effects:
                         Estimate  Std. Error  t value
 (Intercept)              2.28750    0.98144    2.331
 introtypeA2              0.89878    0.73027    1.231
 pictypeS                −0.03159    1.05335   −0.030
 sentencetypeR            2.68007    1.04238    2.571
 pictypeS:sentencetypeR  −3.30079    1.47591   −2.236
```