



Influence of visual and linguistic context on object pronoun processing in a blank screen paradigm

Bachelor's Project Thesis

W. J. van der Rest, s2873672, wesselvdrest@gmail.com,

Supervisors: dr. J.P. Borst & dr. J.C. van Rij-Tange

Abstract: This study aims to investigate whether and how 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 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

The way we humans process sentences can be heavily influenced by slight adjustments in those sentences. How we interpret sentences is a complex process, in which listeners need to combine various types of information such as grammar, lexical meaning, and context information.

This study looks at pronouns in a sentence. In contrast to actual nouns ('Max', 'penguin') pronouns ('he', 'her') can refer to several different referents in the linguistic and non-linguistic discourse. People base their choice of referent on information coming from features such as preceding linguistic discourse, world knowledge and visual context (e.g., Arnold, 1998). Linguistic factors that influence the choice of referent are for example the grammatical role of the referent (a referent to a previous utterance is preferred to be a subject, Chomsky 1981), introduction order of the referent (the first referent is preferred over later referents, Gernsbacher et al. 1988), and recency (referents that are closer to the pronoun are preferred, Gernsbacher et al.

1989).

In the dutch language, object pronouns ('hem' *him*, 'haar' *her*) in simple transitive sentences ('The monkey is photographing him with a photo camera') can never refer back to the subject of the local clause (in this case 'the monkey'). The pronouns always have to refer to another referent in the context. In linguistic theory this restriction in interpretation is known as Principle B of binding theory (Chomsky, 1981). This would suggest that object pronoun processing is mainly influenced by grammatical context, and if linguistic discourse has an influence at all, this effect can be observed later (e.g., Nicol & Swinney, 1989; Clifton, Kennison, Albrecht, 1997; Chow et al., 2014).

In the recent years some studies are suggesting the contrary. For example van Rij (2012) showed that the performance on object pronoun processing tasks can be influenced by visual context and the introduction order of the context. To manipulate the discourse prominence of the referents, the order of introduction of the two referents was swapped. When the agent of the target sentence with the ob-

ject pronoun was introduced first in the linguistic context, the visual context affected pronoun resolution. When the agent was introduced second, it did not. She saw differences in pupil dilation 500-1000ms after the object pronoun onset. This suggests that with a canonical introduction order (when the actor is mentioned first), people have a harder time interpreting the sentence when the pronoun is not congruent with the visual context than when it is congruent (they have to interpret it again). This indicates that linguistic and visual context compete with grammatical principles during object pronoun resolution.

However, pupil dilation can take on average up to 1000ms for an effect to be visible (e.g., Hoeks & Levelt, 1993), which is relatively slow, the latency of pupil dilation also varies a lot between persons, and age is also an increasing factor (Lobato-Rincón, 2014). Although this method is reliable in detecting differences in cognitive processing, it does not have a very high temporal resolution. A way to get more precise information about the timing of the effects of visual and linguistic context is electroencephalography. EEG has a high temporal resolution and the timing of the ERP differences can be related to a long tradition of EEG research in psycholinguistics. In this study the EEG measurement is added to receive a higher temporal resolution than when only pupil dilation is measured. To avoid any eye movements during the processing of the sentences, the experiment is a combination of a picture verification task and the blank screen paradigm (Altmann, 2004). A picture verification task is a task where the participant has to judge if the picture is congruent with an auditory description, in this case the following test sentence. The blank screen paradigm is the name given to the method of removing the visual context before the linguistic context starts. Altmann showed in his study that it is not necessary to show both visual and linguistic context at the same time since people will still move their eyes to the location of the previously shown object even when the object was no longer present. This shows that people react in the same way, with regards to gaze direction, if the visual and linguistic context is presented at the same time than when it's not. If the visual context is presented at the same time of the measurement, the EEG data will show a lot of artifacts due to eye movements. By showing a blank screen during sen-

tence processing, the number of eye movements will be reduced. This method also allows us to distinguish the brain activity caused by sentence processing from the activity caused by the visual context. Thus in contrast to the study by van Rij (2012) this study co-registers EEG data and pupil size. For measuring EEG the blank screen paradigm is used rather than earlier picture verification tasks. The combination of EEG and Pupil size data allows us to compare the results of the EEG data to the earlier found results on pupil dilation (van Rij, 2012).

This study will investigate the question "Do visual and linguistic context (introduction order) have an influence on object pronoun processing?". If this is the case, it is also interesting to investigate what the effects are on pupil dilation, if the same effects are detected in the EEG data and when they precisely occur.

Our hypothesis is that the visual and linguistic discourse context set up an expectation for an interpretation that may be congruent with the pronoun interpretation or not. If the actor is introduced in a canonical introduction order (so introduced first), the visual context will affect pronoun processing. When the actor is introduced second it will not affect pronoun processing. Just as van Rij (2012) found in her research, we expect to see a difference in pupil dilation 500-1000ms after the pronoun onset. The pupil will dilate more when the actor is mentioned first and the pronoun is not congruent to the visual context, than when it is congruent.

We expect to see a higher N400 amplitude when the pronoun interpretation is not congruent to the visual context than when it is congruent (more difficult retrieval, according to Brouwer & Hoeks, 2013; or violating expectations, according to van Berkum et al, 2007). When the mismatch between the visual context and the pronoun interpretation is only detected after grammatical processing, we expect to see later amplitude differences, for example in the P600 window (associated with semantic and pragmatic integrative processes, according to Brouwer & Hoeks, 2013).

2 Method

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

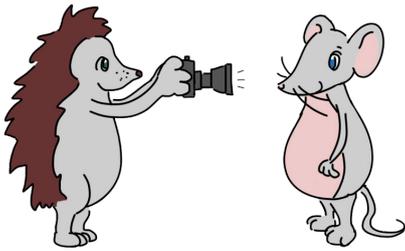


Figure 2.1: An other-oriented image.

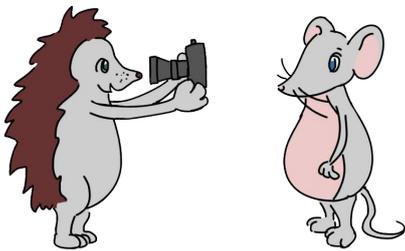


Figure 2.2: A self-oriented image.

By combining a picture verification task with a

blank screen paradigm, we investigated the influences of both visual and linguistic context on object 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 de-

pended 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>+ 100 ms silence. 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 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 ques-

tions, 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.

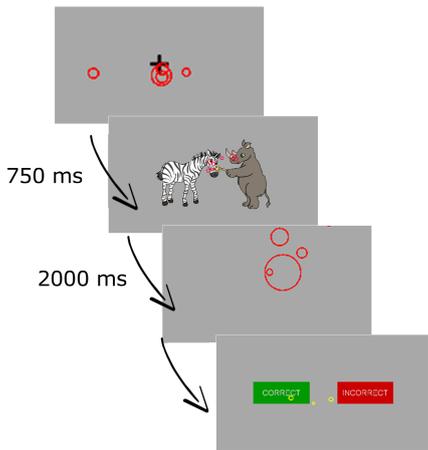


Figure 2.3: Visualization of a trial.

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 investi-

gate 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 incongruity 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.

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

We will first discuss the behavioral data (Subsection 3.1), followed by the data of the pupil dilation (Subsection 3.2) and the EEG data (Subsection 3.3). Finally a comparison between the two will be made. A statistical analysis will be performed on the pupil dilation data and the EEG data to see if any of the conditions do indeed have an influence on the object or reflexive pronoun processing.

3.1 Behavioral data

Figure 3.1 shows the reaction times of the participants on our experiment, Figure 3.2 shows the accuracy. The pronoun trials (‘him’) can be seen on the left side on a pair of bars, and the reflexive trials (‘himself’) on the right side of the pairs. The text ‘Other’ on the left represents the other-oriented pictures, and the text ‘Self’ represents the self-oriented pictures. In both graphs ‘A1’ stands for the actor-first introduction order, and ‘A2’ stands for actor-second introduction order.

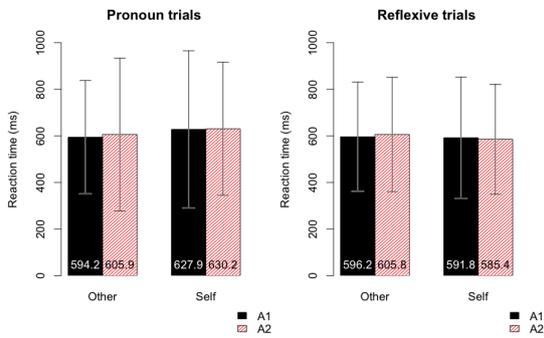


Figure 3.1: The reaction times of the participants.

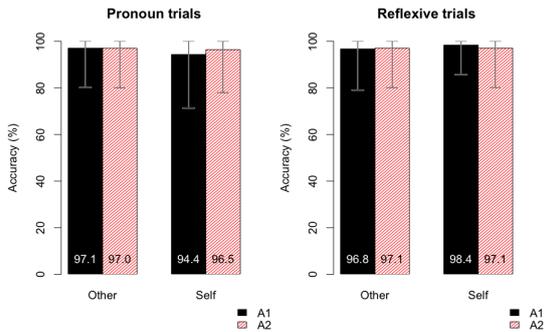


Figure 3.2: The accuracy of the participants.

In Figure 3.1 you see the reaction times of the participant. In both sentence type conditions and both picture type conditions the average reaction times are around 600ms. With these reaction times the participants managed to get an average accuracy of 97% (see Figure 3.2).

There is a small difference in accuracy between

the congruent and incongruent conditions. With the pronoun trials the other-oriented pictures are congruent and the self-oriented pictures are incongruent, the incongruent condition has a lower accuracy percentage. The same holds for the reflexive trials, here the other-oriented pictures are incongruent and the self-oriented pictures are congruent.

3.2 Pupil dilation

Figures 3.3 and 3.4 show the results of the pupil dilation in graphs. Figure 3.3 shows the pronoun test sentence, so the sentence ‘the mouse touched **him** with a spoon’. On the y-axis represents the pupil size. This is an arbitrary unit, chosen by EyeLink, an increase means that the pupil size gets bigger. The x-axis shows the time in ms. the dotted line on $t = 0$ represents the onset of the pronoun.

These two graphs and the graphs in Section 3.3 use the same design to show all conditions. The black lines represent an other-oriented picture (i.e. **congruent** with pronoun test sentences, **incongruent** with reflexive test sentences). The red lines represent a self-oriented picture (i.e. **incongruent** with pronoun test sentences, **congruent** with reflexive test sentences). The legend also shows both solid lines and dashed lines, these show the differences in introduction type. The solid lines represent the actor-first sentences, “*You just saw a mouse and a squirrel*”. The dashed lines represent the actor-second sentences, “*You just saw a squirrel and a mouse*”.

Figure 3.3 shows that all lines follow a similar curve, so there is not much of a difference between them. This would indicate that there is no significant effect of either the picture type or the introduction order.

The graph that show the pupil dilation during reflexive test sentences is shown in Figure 3.4. Here the other-oriented pictures (the black lines) are incongruent, and the self-oriented pictures (the red lines) are **congruent**. The other-oriented actor-first condition (the solid black line) shows a high peak around 1000 ms after pronoun/reflexive onset. This peak is not visible in the other-oriented actor-second condition nor is it visible in both the self-oriented conditions.

Comparing both the reflexive and pronoun test sentence graphs, a big difference is visible between the two graphs, especially on the other-oriented

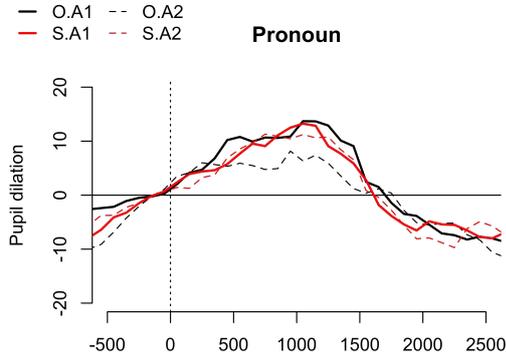


Figure 3.3: Pupil dilation during pronoun test sentences.

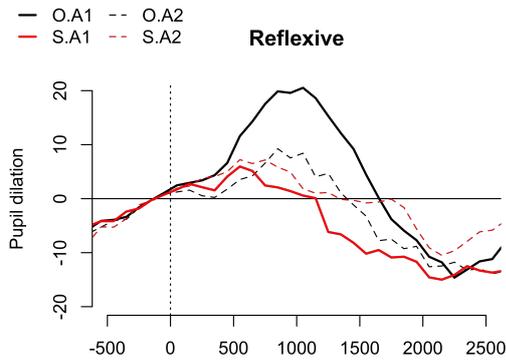


Figure 3.4: Pupil dilation during reflexive test sentence.

condition. This may indicate that picture type and introduction order have an effect on the reflexive test sentences and not on the pronoun test sentences.

To find the best fitting model for the N400 time window a backward-fitting model comparison procedure has been performed on linear mixed effect models. The most complex model, with the three-way interaction shows significantly more variance than the two way interaction model ($X^2(1) = 5.7402, p = 0.016581$). So the most complex model, with the three-way interaction, is the best fitting model, and thus this model is used (see Table 3.1).

Table 3.1 shows the fixed effects in the complex

Model: medianPupil \sim (<i>introtype</i> + <i>pictype</i> + <i>sentencetype</i>) ³ + (1 Subject) + (1 Item)			
	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: sentence-typeR	-18.022	10.116	-1.782
pictypeS: sentence-typeR	-41.064	10.057	-4.083
introtypeA2: pictypeS: sentence-typeR	34.175	14.255	2.397

Table 3.1: Effects on Pupil dilation

model. There is a significant main effect on the sentence type ($\beta = 17.684, SE = 7.146, t = 2.475$ for the reflexive sentences), which indicate that the reflexive sentences elicit a bigger pupil size than the pronoun sentences with an actor first introduction order.

Table 3.1 also shows that the interaction between picture type and sentence type is significant ($\beta = -41.064, SE = 10.057, t = -4.083$ for self-oriented pictures with a reflexive test sentence). This indicates that there is an interaction effect between the two variables. The interaction shows that the effect on pupil dilation is lower when the self-oriented picture is presented together with the reflexive test sentence than with the pronoun test sentence.

Finally the interaction between introduction type, picture type and sentence can also be seen to be significant in Table 3.1 ($\beta = 34.175, SE = 14.255, t = 2.397$ for the actor second introduction order together with self-oriented pictures and a reflexive test sentence). This again shows that the pupil size will be bigger when the actor is introduced second, together with a self-oriented picture and a reflexive test sentence.

Because we see a large difference between the pronoun test sentence and reflexive test sentence

graphs (3.3 and 3.4) we have performed separate analyses on both conditions.

For the pronoun condition the main effects model was used since the two-way interaction did not show any significant deviation from the simplest model ($X^2(1) = 0.0178, p = 0.8939$). The pronoun condition did not show any significance on any condition, the reflexive condition however did show some significance, part of the summary of the analysis is shown in Table 3.2. In this analysis the more complex model, with the two-way interaction, was used since this model showed more significant variance from the simplest model ($X^2(1) = 11.097, p = 0.0008646$).

Model: medianPupil $\sim (introtype + pictype)^2 + (1 Subject) + (1 Item)$			
	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 3.2: Effects of Reflexive sentences on pupil dilation

Table 3.2 shows that all conditions are significant. The introduction type is significant ($\beta = -23.140, SE = 6.957, t = -3.326$ for when the actor is introduced second), which means that the effect on pupil size is smaller when the actor is introduced second during a reflexive test sentence. The picture type also shows significance ($\beta = -32.672, SE = 6.865, t = -4.760$ for a self-oriented picture), this means that the effect on pupil size is smaller during a self-oriented picture with a reflexive test sentence. Finally Table 3.2 also shows that the interaction between introduction type and picture type is significant ($\beta = 32.603, SE = 9.761, t = 3.340$ for when the actor is introduced second combined with a self-oriented picture). This significant interaction shows that the effect on pupil size will be larger during a reflexive test sentence when the actor is introduced second and a self-oriented picture is presented than when it is presented with a other-oriented picture.

3.3 EEG Data

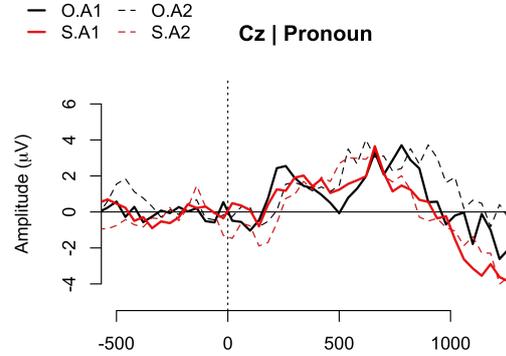


Figure 3.5: EEG data during pronoun test sentences.

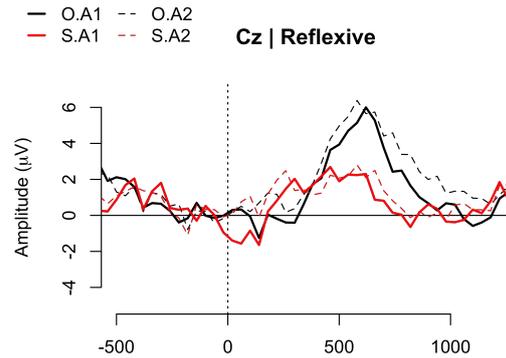


Figure 3.6: EEG data during reflexive test sentences.

For the EEG data we used the Cz electrode: this electrode is located on top and middle of the head. In the Figures 3.5 and 3.6 the Y-axis presents the amplitude in mV, and the X-axis the time in ms. Because EEG data shows an effect faster than pupil dilation the time window here is smaller.

Figure 3.5 shows the baselined EEG data of the pronoun test sentences. Similar to the pupil dilation graph (Figure 3.3 the lines follow a same curve and so no differences between the conditions. This would indicate there is no effect of picture type nor introduction order on the pronoun processing.

In contract with the EEG amplitude during pronoun processing, Figure 3.6 shows a clear congruency effect on the reflexive data. There is a clear difference between the data of the other- and self-oriented pictures, the black lines (so the incongruent condition) have a higher positive amplitude than the red lines around 600 ms after pronoun/reflexive onset. This would indicate that the picture type has an effect on reflexive processing.

We expected to see a peak around 400 ms after pronoun/reflexive onset, representing an N400. A larger peak can be seen around 600 ms after pronoun/reflexive onset, representing a P600. A statistical analysis was performed separately on the N400 time window and the P600 time window (see Tables 3.3 and 3.4).

To find the best fitting model for the N400 time window a backward-fitting model comparison procedure has been performed on linear mixed effect models. The most complex model, with the three-way interaction, is not significantly deviation from the two-way interaction model ($X^2(1) = 0.005, p = 0.9445$). Comparing the two-way interaction model to the main effects model also shows no significant deviation ($X^2(3) = 0.8718, p = 0.8322$), so the main effects model is the best fitting model, and this model is used (see Table 3.3).

Model: $Cz \sim (introtype + pictype + sentencetype) + (1 Subject) + (1 Item)$			
	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 3.3: Effects on N400 in EEG data

Table 3.3 shows no significance in any of the conditions, this would indicate that neither picture type, introduction order, nor sentence type has an effect on object pronoun processing.

To find the best fitting model for the P600 time window another backward-fitting model comparison procedure has been performed on linear mixed effect models. the most complex model, with the three-way interaction, is not significantly deviation from the two-way interaction model ($X^2(1) =$

$0.1227, p = 0.7261$). Comparing the two-way interaction model to the main effects model also shows no significant deviation ($X^2(3) = 5.2345, p = 0.1554$), so from these models the main effects model is the best fitting model.

When we look at the summary of the two-way interaction model we see that the interaction between picture type and sentence type is significant ($\beta = -3.31080, SE = 1.47590, t = -2.243$ for self-oriented picture type and a reflexive test sentence). When we add this interaction to a new model and compare that to the best fitting model we see that there is a significant deviation between the two models ($X^2(1) = 4.9742, p = 0.02573$). So this last model is the best fitting model, and so this model is used (see Table 3.4).

Model: $Cz \sim introtype + pictype + sentencetype + pictype:sentencetype + (1 Subject) + (1 Item)$			
	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

Table 3.4: Effects P600 in EEG data

Table 3.4 shows the results in the time window around 600 ms after pronoun/reflexive onset. The statistical analysis shows that sentence type is significant ($\beta = 2.68007, SE = 1.04238, t = 2.571$ for a reflexive test sentence), this suggests that a higher P600 is elicited when a reflexive test sentence is used. The interaction between the picture type and the sentence type is also significant ($\beta = -3.30079, SE = 1.47591, t = -2.236$ for a self-oriented picture type with a reflexive test sentence). This interaction indicated that a lower P600 effect occurs when a self-oriented picture is used in combination with a reflexive test sentence than when it is used in combination with a pronoun test sentence.

So there are differences between the effects of picture type, introduction order and sentence type on pronoun processing:

- We see no indication of an effect of picture type and introduction order on pronoun processing
- The interaction of picture type and introduction order on reflexive pronoun processing shows that with a canonical introduction order (actor-first) the incongruent picture elicits a larger pupil size than when the actor is introduced second.
- The EEG data shows a larger P600 effect on reflexive processing during an incongruent picture than on a congruent picture.

4 Discussion

This study aimed to answer the question "Do visual and linguistic context (introduction order) have an influence on object pronoun processing?", added by how and when these effects will occur. What our results show is that there is an effect of the visual and linguistic context on the processing of reflexives. In Figure 3.6 a clear difference can be seen around 600 ms after pronoun/reflexive onset, the other-oriented pictures show a stronger P600 effect than the self-oriented pictures on reflexive sentences. This effect cannot be seen in the pupil dilation results though (see Figure 3.4), here only the other-oriented actor first condition shows a high peak around 1000 ms after pronoun/reflexive onset.

We expected to see an effect of the visual and linguistic discourse context setting up an expectation of the interpretation of whether the discourse context is congruent with the pronoun interpretation or not. Our hypothesis states that if the actor is introduced in a canonical introduction order, the visual context will affect pronoun processing. When the actor is introduced second it will not affect pronoun processing.

Clearly our hypotheses are not confirmed, we do not see an effect of the visual and linguistic context on pronoun processing. Both the EEG data and the eye-tracking data show no significant effects during the pronoun test sentences. We did however find significant effects of picture type and introduction order during the reflexive test sentences. In the EEG data we found no significant effect during the N400 time window, but we found the effects in the P600 time window.

The Reflexive test sentence EEG data and eye-tracking data shows that when the actor is introduced first, the pupil dilation will be bigger. These are the exact results we have expected to see on the pronoun processing data. What we did not expect is that there would be a difference between the effect of the introduction order combined with the other-oriented pictures on pronoun/reflexive processing. Figure 3.4 shows a clear difference in the effect of introduction order between the other-oriented conditions. In the EEG data (see Figure 3.6) we see more logical results, that the effect occurs on both actor first and actor second conditions.

One potential explanation is that this is a logical result since "*him*" is more ambiguous than "*himself*", Principle B states that "*himself*" can only refer to the actor, while "*him*" can refer to many other subjects. So where "*himself*" can only have one potential referent, "*him*" can have many others. This makes "*him*" more ambiguous than "*himself*". For this theory we can interpret the P600 peak as a surprise effect, instead of a syntactical violation. So, the surprise is bigger when the reflexive is not congruent. People are not suspecting "*himself*" to refer to the other animal, so they are very surprised when the scene was not congruent. However, since "*him*" is more ambiguous, they are less surprised if the scene is incongruent as it might in fact have referred to the actor.

Since there is a canonical introduction, which people are used to, people are more surprised when the scene turns out to be incongruent. However, when the actor is introduced second, people start to pay more attention as it is not natural to them and are thus less surprised when the scene turns out to be incongruent. Apparently, this effect is only reflected in pupil dilation, but not in EEG.

The P600 peak in Figure 3.6 can also be interpreted as a detection of a mismatch. This would mean we would only see this effect in the incongruent conditions. The eye-tracking data showed that the interaction between the self-oriented picture type and the reflexive test sentence causes the pupil size to get smaller (congruent condition). The EEG data also confirms this since the activity in the Cz electrode decreases when the self-oriented picture is presented together with the reflexive test sentence (congruent condition).

Next to these potential explanations for the differences between our results and the results of the

study by van Rij (2012), another explanation is that there was a difference in how the participants had to perform the experiment.

Most participants had the same complaints about the experiment, they all said it was too long / a dull experiment. Therefore, we decided it would be best to take a closer look at the first two blocks. Some participants also said that it was harder to distinguish the two animals when both animals had the same color, than when there is a clear difference between the animals. Some animals and objects were also thought to be more ambiguous than others, for example the crocodile looking more like a dinosaur. Another complaint was that the self-oriented images were ‘easier’ than the other-oriented images since it is the action is clearer when it is other oriented.

In the results we did not see any significant differences in accuracy or reaction times because of these points though. So, it was not necessary for us to use this in the analysis.

The next step is to continue the analysis and look for the reason why we are getting the contradictory results that we have. Since we added the blank screen paradigm to the experiment there is a possibility this weakens the surprise effect. In the study by van Rij (2012) the visual and linguistic context was presented at the same time, in our study we had to add the blank screen paradigm since there would be more artifacts in the data if the context was presented at the same time. If the visual and linguistic context is presented at the same time, the participant confirms if the linguistic context is congruent to the visual context by looking at the visual context. In our experiment the participant sees the visual context and builds up an expectation of whether the linguistic context should be a reflexive test sentence or a pronoun test sentence. Another explanation could be that we are analyzing the wrong time windows, the effect may already have occurred earlier, or may occur later. Compared to the study by van Rij (2012) our test sentences were formulated in the past tense, while the test sentences in the study by van Rij (2012) were formulated in the present tense. Because the blank screen paradigm was added to our experiment, we chose to use the past tense for the linguistic context because we wanted the sentences to sound as naturally as possible.

To confirm the results of our research the ex-

periment can be conducted with different images and different linguistic context. These images were designed for children, adults may find them less interesting than children.

This study tried to answer the question whether visual and linguistic context (introduction order) have an influence on object pronoun processing. The eye-tracking data shows that when the actor is introduced second and a self-oriented picture is presented, during a reflexive test sentence, the pupil size will become bigger. In the P600 time window the EEG data also shows that the self-oriented picture causes less brain activity in the Cz electrode during a reflexive test sentence condition. So visual and linguistic context does have an effect on reflexive processing, no effects were found on pronoun processing.

5 Acknowledgement

Special thanks to Anouk Hoekstra for letting us record her presenting the linguistic context, Robert Prins and Petra van Berkum for creating the pictures used as visual context, and our first participant Pieter Otte for his patience during the test of the experiment.

References

- G. T. Altmann. Language-mediated eye movements in the absence of a visual world: The ‘blank screen paradigm’. *Cognition*, 2(93):79–87, 2004.
- G. T. Altmann and Y. Kamide. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 3(73):247–264, 1999.
- J.E. Arnold. *Reference form and discourse patterns*. Stanford University. Dept. of Linguistics, 1998.
- Paul Boersma and David Weenink. *Praat: doing phonetics by computer [Computer program]*, 2018. version: 5.4.22 ,URL: <http://www.praat.org/>.
- H. Brouwer and J. C. J. Hoeks. A time and place for language comprehension: mapping the n400 and

- the p600 to a minimal cortical network. *Front Hum Neurosci*, 7(758), 2013.
- Jason E. Albrecht Charles Clifton Jr., Shelia M. Kennison. Reading the words her, his, him: implications for parsing principles based on frequency and on structure. *University of Massachusetts*, 1997.
- N. Chomsky. *Lectures on government and binding*. Studies in generative grammar. Foris Publications, 1981.
- Lewis S. Chow, W. Y. and C. Phillips. Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*, 5(630), 2014.
- A. Delorme and S. Makeig. *EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics.*, 134 edition.
- Hargreaves D. J. Gernsbacher, M. A. and M. Beeman. Building and accessing clausal representations: The advantage of first mention versus the advantage of clause recency. *Journal of Memory and Language*, 28(6):735–755., 1989.
- Morton Ann Gernsbacher and David J Hargreaves. Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27(6):699 – 717, 1988. ISSN 0749-596X. doi: [https://doi.org/10.1016/0749-596X\(88\)90016-2](https://doi.org/10.1016/0749-596X(88)90016-2).
- Bert Hoeks and Willem J. M. Levelt. Pupillary dilation as a measure of attention: a quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*, 25(1):16–26, Mar 1993.
- Rommers J. Huettig, F. and A. S. Meyer. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica*, 2(137):151–171, 2011.
- Luis-Lucio Lobato-Rincón, Maria Del Carmen Cabanillas-Campos, Cristina Bonnin Arias, Eva Chamorro, Antonio Murciano, and Celia ramos sanchez. Pupillary behavior in relation to wavelength and age. 8:221, 04 2014.
- SR Research Ltd. *SR Research Experiment Builder 2.1.140 [Computer software]*. Mississauga, Ontario, Canada, 2017.
- S. Luck. An introduction to the event-related potential technique (cognitive neuroscience). *MIT Press*, 2005.
- J. Nicol and D. Swinney. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18(1):5–19, 1 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/>.
- Zwitserlood P Kooijman V Hagoort P. Van Berkum JJ, Brown CM. Anticipating upcoming words in discourse: evidence from erps and reading times. *J Exp Psychol Learn Mem Cogn.*, 2005.
- J. van Rij. Chapter 6: Event-related pupil dilation tracks integration of contextual and grammatical information. In *Pronoun processing: computational, behavioral, and psychophysiological studies in children and adults.*, pages 115–132. 2012.
- Jacolien van Rij, Bart Hollebrandse, and Petra Hendriks. *Children’s eye gaze reveals their use of discourse context in object pronoun resolution*, pages 267–293. *Linguistische Arbeiten*. De Gruyter, 2016. ISBN 978-3-11-045968-5.