In the Eye of Object Pronoun Processing: Investigating the Influences of Visual and Linguistic Contexts Using Pupil Dilation

Bachelor's Project Thesis

Sho Cremers, s.a.cremers@student.rug.nl,
Supervisors: Dr J.C. van Rij & A. Toth

Abstract: There are two major theories that explain the time course of pronoun processing of object pronouns (e.g., her, him). The initial-filter account suggests grammatical principles restrict the interpretation of object pronouns. The competing-constraints account, on the other hand, suggests other sources of information compete with grammatical constraints. It can be evidence for the competing-constraints account if non-grammatical information influences the interpretation of object pronouns early after the onset of the pronoun. A past study has successfully found evidence for the influence of visual and linguistic contexts, but the follow-up studies have yielded contradictory results. This study aimed to replicate the results of the initial study by investigating the influence of linguistic and visual contexts during the processing of object pronouns with native Dutch speakers. The linguistic and visual contexts were manipulated to construct a within-subject design, as well as a between-group design for the difference in the time delay between visual and linguistic context. Unfortunately, not enough data was gathered to study the effect of the time delay between the visual and linguistic context. However, evidence for the influence of linguistic context on object pronoun processing was found, supporting the claims of the competing-constraints account.

1 Introduction

"John reminded him to drink water."

Simply given this sentence, you will know that *"him"* refers to someone unknown from the context and that it cannot refer to *John* even though only *John* is mentioned in the sentence. In languages like Dutch and English, this restriction on object pronouns (e.g. her, him, it), such that they cannot refer to the subject of the sentence, is known as Principle B of Binding Theory (Chomsky, 1981).

Grammatical rules such as Principle B of Binding Theory play an important role in the interpretation of object pronouns. Some researchers believe that they are the primary factor during object pronoun resolution. The initial-filter account suggests that we first eliminate possible referents that violate grammatical rules like Principle B of Binding Theory (Chow et al., 2014; Clifton et al., 1997; Nicol and Swinney, 1989). Only then, can other sources such as linguistic and visual information influence pronoun resolution.

On the other hand, the competing-constraints account suggests that other sources of information such as linguistic information compete with grammatical principles during object pronoun resolution (Badecker and Straub, 2002; Kennison, 2003). The supporters of this account believe that the identification of the correct referent can be sped-up by allowing non-grammatical information to take part in pronoun resolution from an earlier stage. This is because it would allow for the identification of potential referents before all the details for grammatical rules are gathered.

A study conducted by van Rij (2012) reports results that support the competing-constraints account. The study found influences of both linguistic and visual information on object pronoun processing. In her study, native Dutch speakers were asked to answer if an image matched the auditory descrip-

tion. Participants were presented with a cartoon image with two animals in which one of the animals was either performing an action towards the other animal or towards itself. Two prerecorded sentences were presented 500 ms later, while the image remained on the screen. The first sentence introduced the two animals. The introduction order was manipulated so that there are trials in which the actor was introduced first, or the actor was introduced last. The second sentence described the image with an object pronoun. The congruency of the image and the interpretation of the object pronoun was manipulated by changing the image—that is, if an image of an animal performing an action to the other animal was presented it was congruent with the second sentence, whereas if an image of an animal performing an action to itself, it was incongruent with the pronoun. Pupil dilation was used as the measurement of cognitive load, thus a pupil of the participant was monitored during the experiment. Effects of both introduction order and congruency were found 500-1000 ms after the onset of the object pronoun. When the actor was introduced first, the incongruent condition elicited a larger pupil dilation compared to the congruent condition. On the other hand, when the actor was introduced last, the congruency did not have a strong influence on the pupil dilation. Furthermore, the actor-second introduction elicited a larger pupil dilation compared to the actor-first introduction for the congruent condition, but the introduction order had the opposite effect for the incongruent condition.

The introduction order's influence on the object pronoun resolution can be explained by first-mention bias (Gernsbacher and Hargreaves, 1988). The initially mentioned referent is expected to have a more prominent role than later mentioned referents. Hence the results can be interpreted as follows. When the actor is introduced first in the congruent items, the subject of the next sentence is expected to be the actor. In this case, the expectation is correct and hence requires less effort at the onset of the object pronoun. However, if the actor is introduced last in the congruent items, there is conflicting information, thus the listener is forced to wait until the onset of the pronoun to interpret. If the image is incongruent, the actor-first introduction creates a false expectation, and hence a larger surprise is observed than the actor-second introduction. The results of the study by van Rij (2012) suggest both linguistic and visual contexts compete with the grammatical principles during the interpretation of the object pronoun, supporting the competing-constraints account. If the initial-filter account was true, manipulating linguistic and visual contexts would not show immediate effects during pronoun processing. Instead, they would only be observable later in the time course of pronoun processing.

A pupil is mainly known to dilate in response to brightness, but it has also been shown to dilate with larger cognitive load (Mathôt, 2018). This finding has also shown in linguistic research, in which tasks with larger cognitive load showed larger pupil dilation (Engelhardt et al., 2010; Zellin et al., 2011), and hence pupil dilation was used as the measurement of the cognitive load in the study by van Rij (2012).

Follow-up studies, however, have not always had consistent results with van Rij (2012). A study by Verhoeven (2018) follows a similar structure as van Rij (2012) with minor differences in the procedure. The prerecorded sentences were presented 2000 ms after the picture appeared on the screen. Also, the picture disappeared when the sentences started playing. In addition to monitoring pupil size, the EEG signal was recorded during the experiment. The peak of pupil dilation due to the increase in mental effort takes approximately 1000 ms, whereas EEG can provide more time-precise information on mental effort. The results did not show significant effects of introduction order and congruency on object pronoun processing, in contrast to van Rij (2012). Instead, an interaction between introduction order and congruency on reflexive pronouns (herself, himself) was found.

Another follow-up study by Marinov (2019) used the visual and auditory stimuli from Verhoeven (2018), where again only pupil size was monitored. Similarly to the study by Verhoeven, the prerecorded sentences were presented 2000 ms after the onset of the image. However, the main difference between the two studies was that the image remained on the screen while the prerecorded sentences were being played. The results suggested that both the introduction order and the congruency influenced object pronoun processing. He also found an interaction effect between them. For the reflexive pronouns, the introduction order and the interaction had significant influence, while the con-

gruency may have been significant.

The purpose of the current study was to investigate the conflicting results of the previous studies (van Rij, 2012; Verhoeven, 2018; Marinov, 2019) on online object pronoun processing. This study was structured similarly to van Rij (2012). The participants were first presented with an image on the screen. Then, two sentences in Dutch were presented aurally while the image remained on the screen. The first sentence introduced the two animals in the image. The second sentence correctly or incorrectly described the image. Participants were then asked to answer if the image matched the description of the image in the second sentence. In contrast to the previous studies, the sentences were created using a speech synthesizer. Also, the first sentence (introduction) had a pause after the first referent, but the second sentence (context) did not have any pause. this is in contrast with van Rij (2012), which had no pauses in either of sentences and Verhoeven (2018) and Marinov (2019), which had pauses on both sentences. The choices of using a speech synthesizer as well as using minimal amount of pauses were to make the audio natural as possible while keeping the sentences from other trials consistent. Filler items included either a reflexive pronoun or a noun instead of an object pronoun, which also differed from van Rij (2012), which used reflexive pronouns and intransitive sentences and Verhoeven (2018) and Marinov (2019), which used only reflexive pronouns as fillers. In addition to the introduction order and the congruency, the time between the onset of the image and the onset of the first sentence was manipulated. One group of participants had a 500 ms delay between the visual and auditory stimuli, whereas the other group had a 2500 ms delay. The pupil size was monitored as a measurement of cognitive load during the experiment.

The current study aimed to answer if visual and linguistic contexts influence pronoun processing of object pronouns when the auditory stimuli are presented shortly after the visual stimulus. If an effect is found:

- Are there relationships between visual and linguistic contexts?
- When can we observe the effect?
- Does the timing of these contexts influence the effect?

We hypothesized that the interpretation of object pronouns would be influenced by visual and linguistic contexts. That is, the introduction order and the image would create an expectation about the pronoun. For the congruent condition, we expected actor-second introduction order, which was an unusual order of introduction in a typical discourse, to elicit larger pupil dilation than when the actor was introduced first. On the other hand, we expected the actor-first introduction order to elicit larger pupil dilation compared to the actor-second introduction for the incongruent condition. We also expected the effect of linguistic and visual contexts to peak at the onset of the object pronoun. In general, we expected the results to be in agreement with van Rij (2012) and Marinov (2019).

Studies by Verhoeven (2018) and Marinov (2019) had conflicting results, we believed the reason was the duration of the visual stimulus on the screen before participants were presented with auditory stimuli. If the audio is played shortly after the image is presented, we expected to have stronger effects of linguistic and visual contexts compared to when there is more time between the presentation of the image and the onset of audio. The longer the time between the onsets of the two stimuli, the more time there is to organize visual information, and hence it would be less affected by following unexpected or contradicting information.

We also expected to have more reliable data than Verhoeven (2018) and Marinov (2019) due to using synthesized speech and having fewer pauses for auditory stimuli. Participants would spend less mental effort for the comprehension of the sentences compared to the previous studies, and hence the change in cognitive load would be a better indication of influences on pronoun processing. Furthermore, having a larger variety of filler sentences would reduce the chance of participants building expectations for the trial as well as getting bored, and thus decreasing the automatic responses. The automatic responses require less effort (Ferreira et al., 2006), and hence the data on the cognitive load during pronoun processing would not be very meaningful if many trials were answered effortlessly. This was important since it was a long experiment that took approximately one hour.

2 Methods

2.1 Participants

A total of 6 native Dutch-speaking subjects participated in the experiment, of whom 4 were male and 2 were female. The mean age was 22 years old ranging from 18 to 24. Participants were recruited by distributing fliers around the University campus, sending e-mails in the department of Artificial Intelligence and Human-Machine Communication, as well as posting advertisements on social media. All participants signed the informed consent form before the experiment. The experiment took approximately 1 hour and the participants received compensation of 8 euros.

2.2 Design

The design of the current study was strongly influenced by the visual world paradigm, which is often used in studying the effect of visual and linguistic information (Huettig et al., 2011). Participants were asked to do a verification task, in which if the image correctly corresponds to the auditory information. It was a $2 \times 2 \times 2$ design with the following conditions:

- Congruency (congruent or incongruent)
- Introduction order (actor-first or actor-second)
- Time between image onset and audio onset (500 ms or 2500 ms)

The congruency and the introduction order was tested within-subjects, and the time condition was tested between-subjects.

In the experiment, two types of pictures were used. Both types of pictures contained two animals. One was other-oriented pictures (Figure 2.1), in which one animal performs an action on the other animal. The other type of picture was self-oriented pictures (Figure 2.2), in which an animal performs an action to itself. Trials in which other-oriented pictures were used were considered the congruent condition and self-oriented image as incongruent with the object pronoun. Eighty unique images were used, which consisted of 40 pairs of a picture with variants of other- and self-oriented pictures.

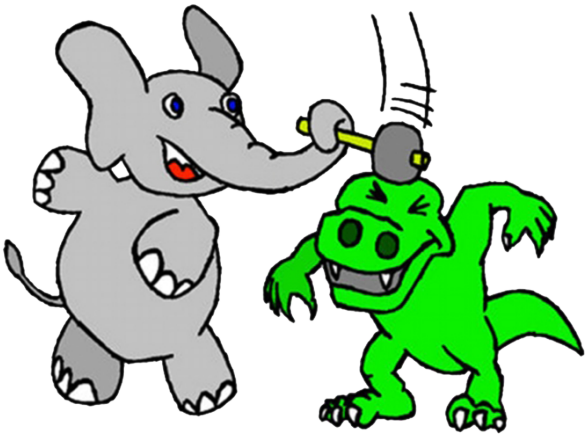

Figure 2.1: An example of other-oriented pictures. Trials with other-oriented images were considered congruent condition.

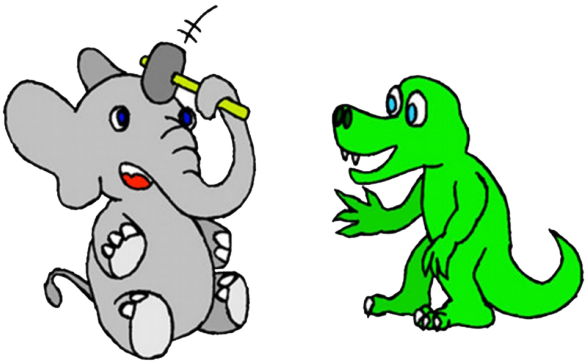

Figure 2.2: An example of self-oriented pictures. Trials with self-oriented images were considered incongruent condition.

The introduction order is related to the introduction sentence, which was presented after the picture. In the actor-first (A1) condition, the actor was introduced first (*"Hier zie je een olifant en een krokodil."*, Here you see an elephant and a crocodile.), whereas in the actor-second (A2) condition the actor was introduced last (*"Hier zie je een krokodil en een olifant."*, Here you see a crocodile and an elephant.).

Image duration was manipulated to test the influence of timing of visual and linguistic context. One group of the participants was provided with the first sentence (introduction sentence) 500 ms after the picture appeared on the screen. The other group heard the first sentence 2500 ms after the

picture.

One block included 40 items, and the 500 ms group of the time condition completed 4 blocks, whereas the 2500ms group completed 3 blocks. The difference in the number of blocks between the two time conditions was to have both groups take approximately the same time to complete the experiment since a trial of the 2500 ms group was 2 seconds longer than a trial of the 500 ms group. For each subject, a unique list of 160 items was created. Each image was used twice, one of the times it was presented in block 1 or 2, and the second time in block 3 or 4. Images were mirrored with a chance of 50% in blocks 1 and 2. If the same image used in block 1 or 2 was non-mirrored, then the image was mirrored for block 3 or 4, and vice versa. The order of the items in each block was randomized to avoid the participants' expectations for the next item.

2.3 Materials/Stimuli

The presentation of visual and auditory stimuli was programmed in Experiment Builder (SR Research, 2019).

The pictures used in Verhoeven (2018) and Marinov (2019) were presented as a visual stimuli. Images were scaled to have the larger side (length or width depending on the image) to have 500 pixels. The size of the monitor was 1920×1080 pixels and the picture was placed in the center with a gray background.

Two sentences were used to describe each picture. These sentences were created using Google (2019)'s Text-to-Speech API with the voice name 'nl-NL-Wavenet-A', speed of 85%, and unmodified pitch.

Two introduction sentences were created for each picture. The structure of these sentences is: "Hier zie je" (Here you see) + <referent> + 75 ms pause + "en" (and) + <referent>. The introduction order was manipulated by switching referents. For the A1 condition, the actor was the first referent, whereas the actor was the second referent in A2 condition.

Test sentences were played after the introduction sentence had ended. These test sentences contained the referring expression, which included either an object pronoun, reflexive pronoun, or noun. In total, 4 test sentences were created for each picture. The test sentences were structured as follows: <Actor> + <verb> + <pronoun/reflexive/noun> + <prepositional phrase>. Filler items included a reflexive pronoun or a noun instead of an object pronoun. In each block, 50% of the trials contained the object pronouns and 50% of the trials used the fillers (25% reflexive, 25% noun). For the reflexive condition, the trial was considered congruent if the image used was self-oriented, whereas it was incongruent if an other-oriented image was used. In the case of the noun condition, it was always referred to the non-actor, and sentences with a correct prepositional phrase and an incorrect prepositional phrase were created.

The answer screen was presented 1500 ms after the test sentence had ended playing. The texts "Correct" in a green box and "Incorrect" in a red box appeared on the left side and the right side of the screen. The positions of correct/incorrect boxes were randomly chosen to prevent the participants from preparing to answer before the screen appears. The "Q" key was used to indicate the answer of the left side and "P" key for the answer of the right side.

2.4 Procedure

Participants were informed not to wear mascara, glasses, nor contact lenses in advance since it could affect the recording of the pupil. Before the participant arrived at the eye-tracking lab, the humidifier was turned on to reduce the number of blinks during the experiment.

When the participant arrived, they were first told to read and sign the informed consent form. Once the form was signed, they were asked to answer a small questionnaire that was used to obtain demographic data.

Participants were seated on a non-adjustable chair and the table of the headset was adjusted. Then, the chin-rest of the head stabilizer was adjusted, in which the participants were able to sit comfortably. The eye-tracker was placed in front of the screen. The keyboard was placed between the head stabilizer and the eye-tracker, which was later used by participants to give the answers.

When the participants were set up, they were presented with written instructions on the screen. When they were done reading the instructions, a 9-point calibration was performed. After the calibration, 3 practice trials were presented. All the participants were presented with the same practice

trials, but the time condition was applied. The images used in the practice trials were not used in the experiment. For the first 5 subjects, the pupil of the left eye was measured and pupil of the right eye for 1 subject during the experiment using Eyelink Portable Duo (SR Research). The reason for the change in the eye of measurement was because we learned the right eye is more commonly the dominant eye of a person (Porac and Coren, 1976) after starting the conduction of the experiment. Pupils were monitored with the sample rate of 1000 Hz and illumination was set to 75%.

An overview of the trial sequence is presented in Figure 2.3. Each trial started with a drift correction: a fixation appeared in the center of the screen, and the subjects were told to look at the point. Once it was validated, a blank grey screen was displayed for 250 ms until the picture appeared on the screen. After 500 ms or 2500 ms (depending on the condition) the introduction sentence was presented, then immediately after, the test sentence was presented while the picture was still displayed on the screen. The answer screen appeared on the screen 1500 ms after the test sentence has ended. The subject had 5000 ms to press either "Q" or "P" Key to answer. If the subject did not answer within 5000 ms, it proceeded to the next trial.

After the 3 practice trials were done, the subjects were allowed to ask questions if they were unclear about the experiment. When everything was clear, they began their first block.

Each block started with calibration and validation. The block consisted of 40 items. After each block, a short break was given. Subjects in the 500 ms condition were presented with 4 blocks (total of 160 trials), and the subjects in the 2500ms condition were presented with 3 blocks (total of 120 trials).

After the experiment had ended, the participants were asked for their bank details for the compensation.

2.5 Analyses

The pupil dilation data were pre-processed using a script from Felix Boie in R (R Core Team, 2018) version 3.6.0. In the script, blinks were removed from data. Only the trials of the object pronoun condition were used for the analyses, and hence trials from reflexive pronoun and noun conditions

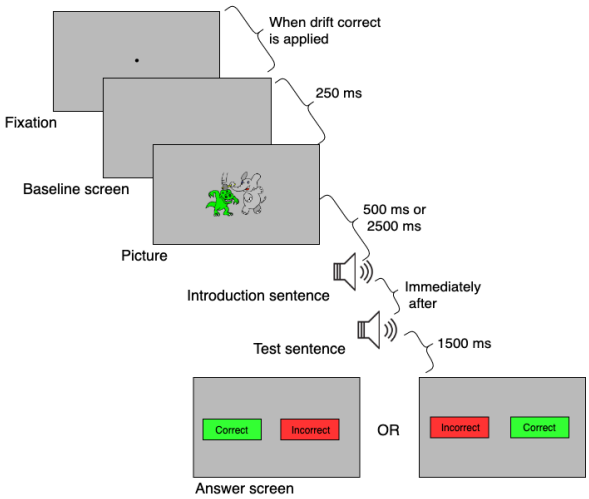

Figure 2.3: An overview of the trial sequence.

were removed. If more than 30% of the data from the trial was missing (NA), the whole trial was excluded from the data for analyses. Also, if 80% or more of the data during the baselining was missing, the trial was removed. In total, only 2 out of 400 object pronoun trials were removed. Then, the data were downsampled to 50 Hz from 1000 Hz.

Statistical analyses were performed on the pupillometry data using the linear mixed-effects model (LME) (Winter, 2013) in order to take into account the subjects' differences of pupil dilation, as well as the influence of each item on pupil dilation. The data was baselined on 200-0 ms before the onset of the test sentence. The analyses were done in R (R Core Team, 2018) version 3.6.0, using the package "lme4" (Bates et al., 2015) version 1.1.21.

In total, two points in the time course of a trial were analyzed. The first analysis was performed on the onset of the test sentence. The data was aligned with the onset of the test sentence, and the baselined mean pupil size between 750 ms and 1250 ms after the onset of the test sound was analyzed since it is known that the pupil dilation peaks approximately 1000 ms after the stimulus that triggers the pupil to dilate (Hoeks and Levelt, 1993).

The second analysis was performed at the onset of the object pronoun. The data was aligned with the onset of the object pronoun, and the mean pupil size between 750 ms and 1250 ms after the onset of the test sound was analyzed.

The onsets of object pronoun were determined from the test sentences using an R script R Core Team (2018) provided by Jacolien van Rij. Another set of test sentences were synthesized, which were the same as the original ones, but it contained a 75 ms pause before the object pronoun. The R script was able to find the starting time of the pause, thus estimating the onsets of the object pronoun in the test sentences.

3 Results

The data were pre-processed as explained in the analysis section and a statistical analysis was performed in order to determine if there were effects of the introduction order and the congruency on pupil dilation during object pronoun resolution for the 500 ms group. The data of the 2500 ms group as well as the effect of change in time was not analyzed due to having only one participant in the 2500 ms group.

3.1 Accuracy

During the experiment, the accuracy of the participants' answers was recorded. On average, the accuracy of trials of object pronouns was high ($M_{accuracy} = 98.9\%$), and Table 3.1 suggests that conditions did not affect the accuracy since all the conditions have high accuracy.

3.2 Pupil dilation

In this subsection, the analysis of pupil dilation is discussed. Analysis of the onset of the test sentence and the onset of object pronoun was performed using Linear mixed-effect models (LME).

3.2.1 Onset of the test sentence

Figure 3.1 shows the results of pupil dilation during the pronoun trials for all the combinations of congruency and introduction type conditions. The x-axis represents the time in milliseconds from the onset of the test sentence, and the y-axis represents the pupil size in arbitrary units given by the eye-tracker after baseline. The data was baselined on 200-0 ms before the onset of the test sentence and aligned on the onset of the test sentence. The actor-second, congruent (A2_V) condition has a larger

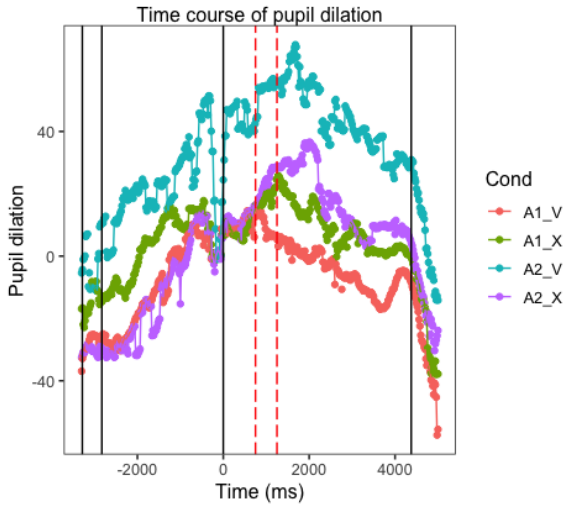

Figure 3.1: Pupil dilation during the trial for each condition. The trials were aligned on the onset of the test sentence. The vertical line on far left (-3283 ms) indicates the average time of onset of the picture. The second line from the left (-2828 ms) indicates the average onset of the introduction sentence. The third line from the left (0 ms) indicates the onset of the test sentence. The far-right vertical line (4377 ms) indicates the average onset of the answer screen. The data between the 2 dashed red lines (750 ms and 1250 ms) were aggregated and further analyzed.

pupil size throughout the trial compared to the other conditions. The actor-first, congruent (A1_V) condition seems to have a smaller pupil size than other conditions for most of the time.

Statistical analysis was further performed on the aggregated data to investigate the effect of the introduction order and the congruency using LME. Chi-squared test revealed that the random effects of both subjects and items did not show significant effects (Subjects: $\chi^2(7) = 0$; $p = 1$, Items: $\chi^2(7) = 3.0252$; $p = 0.82$). Since no significant random effect was found, linear regression models (Navarro, 2018) were used for the following analysis instead of LME.

Backward-fitting model comparison procedure was performed on 2 linear models: a model containing the main effects of introduction type and congruency, and a model containing the same main effects as well as the interaction of the main effects.

Table 3.1: The accuracy of object pronoun trials in 500 ms group for every conditions.

Introduction Order	Congruency	Total items	Correct items	Accuracy
Actor-first	Congruent	100	97	97.0%
Actor-first	Incongruent	100	99	99.0%
Actor-second	Congruent	100	100	100%
Actor-second	Incongruent	99	99	100%

Table 3.2: The estimates of the main effects on pupil dilation of the best fitting model for the onset of the test sentence.

Model: meanPupil ~ introtype			
	Estimate	Std. Error	t value
(Intercept)	13.609	8.767	1.552
introtypeA2	24.633	12.430	1.982

The model comparison did not indicate significant interaction between introduction type and congruency ($F(1, 394) = 2.4324$; $p = 0.12$).

The F test indicated that there may be an effect of introduction order on pupil size ($F(1, 395) = 3.9242$; $p = 0.048$), but not congruency ($F(1, 395) = 0.6838$; $p = 0.41$). Table 3.2 shows the estimates of the main effects of the best-fitted model. It can be interpreted that when the actor is introduced second, the pupil is dilated by 24.6 arbitrary units more than when the actor is introduced first.

3.2.2 Onset of the object pronoun

Figure 3.2 shows the results of pupil dilation during the pronoun trials for all the combinations of congruency and introduction type conditions. The x-axis represents the time in milliseconds, and the y-axis represents the pupil size in arbitrary units given by the eye-tracker after baseline. The data was baselined on 200-0 ms before the onset of the test sentence and aligned on the onset of the object pronoun during the test sentence. The A2_V condition has a larger pupil size compared to the other conditions. The A1_V condition seems to have a smaller pupil size than other conditions. For all conditions, the peak of pupil dilation is in or near the time window of 750 ms and 1250 ms after the onset of object pronoun, suggesting that the onset of object pronoun is the peak of cognitive load during the trial.

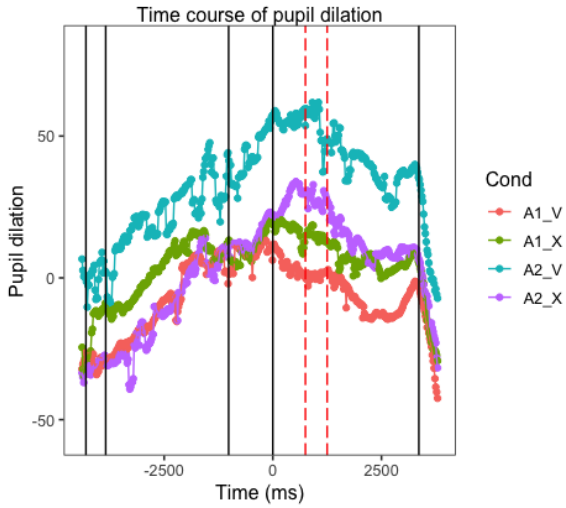

Figure 3.2: Pupil dilation during the trial for each condition. The trials were aligned on the onset of the object pronoun. The vertical line on far left (-4300 ms) indicates the average time of onset of the picture. The second line from the left (-3845 ms) indicates the average onset of the introduction sentence. The third line from the left (-1017 ms) indicates the average time of the onset of the test sentence. The fourth line from the left (0 ms) indicates the onset of the object pronoun. The far-right vertical line (3360 ms) indicates the average onset of the answer screen. The data between the 2 dashed red lines (750 ms and 1250 ms) were aggregated and further analyzed.

Statistical analysis was further performed on the aggregated data to investigate the effect of the introduction order and the congruency. Once again, inclusion of random effects did not indicate an improvement (Subjects: $\chi^2(7) = 0$; $p = 1$, Items: $\chi^2(7) = 2.4491$; $p = 0.12$), and hence the following analysis was done by using linear regression models.

The F test indicated that there is no interac-

Table 3.3: The estimates of the main effects on pupil dilation of the best fitting model for the onset of the object pronoun.

Model: meanPupil $\sim$ introtype			
	Estimate	Std. Error	t value
(Intercept)	6.340	9.575	0.662
introtypeA2	35.645	13.592	2.622

tion between introduction type and congruency ($F(1, 393) = 1.7719$; $p = 0.18$).

Congruency was not a significant indicator for the pupil dilation ($F(1, 394) = 0.0772$; $p = 0.78$), but there was effect of introduction order ($F(1, 394) = 6.8646$; $p < 0.01$) on pupil size. Table 3.3 shows the estimates of the main effects of the best-fitted model. It can be interpreted that when the actor is introduced second, the pupil is dilated by 35.6 arbitrary units more than when the actor is introduced first.

4 Discussion

The present study was conducted in order to investigate the influence of linguistic and visual context on pupil dilation during object pronoun processing of native Dutch speakers. More specifically, the present study was aimed to answer the following questions:

- Are there relationships between visual and linguistic contexts?
- When can we observe the effect?
- Does the timing of these contexts influence the effect?

The influence of visual context was determined by manipulating the congruency of the image to the audio sentences. The influence of the linguistic context was determined by manipulating the introduction order of the animals in the picture. The mental effort during the object pronoun resolution was assessed to test the influence of visual and linguistic contexts by measuring pupil dilation.

The results of the current study suggest that the congruency of the image did not influence the pupil dilation. However, weak evidence was found to support the influence of the introduction order on the pupil size at the onset of the test sentence. Stronger evidence of the effect of the introduction order was found on the onset of the object pronoun. For both cases, it elicited larger pupil dilation for the actor-second introduction compared to the actor-first introduction. The effect of the introduction order also seems to be larger at the onset of the object pronoun than at the onset of the test sentence. There was not enough evidence to support the interaction between the introduction order and the congruency.

Due to the low number of participants, only the data of participants in the 500 ms condition were analyzed, and not the data from 2500 ms. Hence, the current study was not able to answer the effect of the timing of visual and linguistic contexts during pupil dilation.

The present study suggests that linguistic context plays an important role in object pronoun processing. It elicited a larger pupil size when the actor was introduced last compared to the actor-first introduction. This can be interpreted that when the two characters are introduced, listeners expected the first mentioned referent to be the subject in the test sentence and hence the object pronoun would refer to the later mentioned referent. When the actor is introduced later, the later mentioned referent is first chosen as the possible interpretation of the object pronoun. However, as they find out that the referent is the actor of the next sentence, it contradicts with their initial belief and hence increases cognitive load.

Overall, the results were only partially in agreement with van Rij (2012) and Marinov (2019). All three of the studies found the effect of the introduction order on the pupil dilation. However, the effect of congruency and the interaction between the introduction order and the congruency was also observed in van Rij (2012) and Marinov (2019), whereas the current study did not.

Even with the differences in the results, the current study supports the competing-constraints account like the two previous studies (van Rij, 2012; Marinov, 2019). The effect of linguistic context on pupil dilation during the processing of object pronouns can be considered as evidence for the view that the grammatical principles and linguistic context compete during the processing of object pronouns. The fact that stronger evidence of the influence of linguistic context is observed at the onset of the object pronoun compared to the onset of the

test sentence also suggests that the increase in cognitive load comes from pronoun processing, and not just that an unusual introduction surprised the listener.

It was interesting that we did not observe significant random effects of items and especially subjects. The main reason random effects were not observed is thought to be because the baselining process removed the majority of the differences in the initial pupil size.

It is also important to keep in mind that the current study had data for only 5 subjects. Hence, it cannot be concluded that congruency did not affect the object pronoun resolution. This is also the case with the interaction between the two variables.

The use of synthetic speech was a different aspect of the design in comparison to previous studies (van Rij, 2012; Verhoeven, 2018; Marinov, 2019), but it should not have any influences on understanding the sentences. Participants claimed that the Dutch speeches did sound natural, with the exception of the pronunciation of some of the words (e.g. the word *"pinguïn"* had a stronger emphasis than usual), but even with the mispronunciation, the subjects were able to comprehend them.

A minor problem during the experiment was the calibration process during the setup of the eye-tracker. The top corner of the same side of the measuring eye (top-left corner of the screen if the left eye was being measured, and top-right corner if the right eye was being measured) had difficulty capturing the gaze, possibly because of having a wide monitor. However, this should not be an issue in the current study, since the visual stimuli were always presented in the center of the screen and never in the corner.

Although it was unfortunate to have a small number of participants, strong evidence of the influence of introduction order on object pronoun processing from such a small sample size suggests the importance of linguistic context. The results of the present study support the idea of non-grammatical information playing an important role in the interpretation of object pronouns.

5 Acknowledgements

First and foremost, I would like to thank my supervisors Jacolien van Rij and Abby Toth for guidance throughout the project. I would also like to thank Felix Boie for providing us with scripts to help preprocess the pupil dilation data. Last but not least, a big thank you to Kenichi Furusawa for his great partnership during the project.

References

Badecker, W. and Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4):748–769.

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1).

Chomsky, N. (1981). *Lectures on government and binding : The pisa lectures*. Dordrecht: Foris Publications.

Chow, W.-Y., Lewis, S., and Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*, 5:630.

Clifton, C., Kennison, S. M., and Albrecht, J. E. (1997). Reading the WordsHer, his, him:implications for parsing principles based on frequency and on structure. *Journal of Memory and Language*, 36(2):276–292.

Engelhardt, P., Ferreira, F., and Patsenko, E. (2010). Pupillometry reveals processing load during spoken language comprehension. *Quarterly Journal of Experimental Psychology*, 63(4):639–645.

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., and Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of personality and social psychology*, 91(5):797–813.

Gernsbacher, M. A. and Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27(6):699–717.

Google (2019). Cloud Text-to-Speech. `https://cloud.google.com/text-to-speech/`.

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

Huettig, F., Rommers, J., and Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2):151–171.

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

Marinov, B. (2019). Resolving the pronoun problem: Investigating the influence of linguistic and visual contexts on object pronoun processing.

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, 1(1):16–16.

Navarro, D. (2018). *Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6)*, chapter 15: Linear regression. University of Adelaide. `https://learningstatisticswithr.com/lsr-0.6.pdf`.

Nicol, J. and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18(1):5–19.

Porac, C. and Coren, S. (1976). The dominant eye. *Psychological Bulletin*, 83(5):880–897.

R Core Team (2018). R: A language and environment for statistical computing. `https://www.R-project.org/`.

SR Research (2019). SR Research Experiment Builder [Computer software].

van Rij, J. (2012). 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. Rijkuniversiteit Groningen.

Verhoeven, R. (2018). Influence of visual and linguistic context on object pronoun processing: EEG and eye-tracking provide new information.

Winter, B. (2013). Linear models and linear mixed effects models in r with linguistic applications. arXiv:1308.5499. `http://arxiv.org/pdf/1308.5499.pdf`.

Zellin, M., Pannekamp, A., Toepel, U., and Van der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology*, 81(3):133–141.