university of groningen / faculty of science and engineering / artificial intelligence

The role of "er" and "waar" in creating long-distance dependencies in Dutch: an eye-tracking study

Bachelor's Project Thesis

Jules Roboz, s4691091, j.roboz@student.rug.nl
Supervisor: Dr. S.M. (Stephen) Jones, PhD

Abstract: Using an eye-tracking experiment, this study aims to explore the relationship between reading speed and long-distance dependencies (LDDs) formed between the words "er" and "waar" and prepositional object gaps in Dutch. In part, this experiment aims to expand the findings of Stowe (1986) to Dutch, which demonstrated that English WH-constructions form LDDs which lead to slower reading speed at an object gap location. Additionally, it presents another potential candidate for LDD formation unique to Dutch: the word "er". Reading speed was measured and compared across three conditions: a baseline sentence, a WH-construction in the form of a "waar" sentence, and an "er" sentence creating an LDD across separate sentences. Results were analysed using a statistical model which showed no significant effect on reading time across any condition. The methodology of the experiment as well as the underlying hypotheses surrounding gap search procedures in Dutch are discussed.

1 Introduction

1.1 Psycholinguistic background

One of the fundamental questions in human language processing is the matter of whether the process functions in a top-down or a bottom-up direction. When reading a sentence, is the processing driven by an existing prediction for what the structure of the sentence is, or is the internal understanding of sentence structure built up along the way, with each piece processed on its own?

One area of research in this topic is how the brain processes so-called "semantic gaps". A gap occurs when a certain phrase is expected at a certain position in the sentence based on typical sentence structure expectations, but no such phrase occurs, as the expected semantic role has been filled by another constituent, named the "filler". How this discrepancy is resolved during lexical processing is a key question of interest. There are many competing theories, one notable one being the theory of filler-driven parsing. Frazier & d'Arcais (1989) explore how readers recognise and fill in a gap left behind by a prepositional object in Dutch. They argue that during processing, an active filler strategy is employed, wherein potential fillers to the gap are immediately identified as they come up in the text, and parsing is driven by the current expectation of what the filler will be. This refutes previous bottom-up theories (Jackendoff & Culicover, 1971; Wanner & Maratsos, 1978; Fodor, 1978, 1979, as cited in Frazier & d'Arcais, 1989) that a gap is postulated only when parsing local phrases next to the gap which indicate its presence (in other words, processing is gap-driven, not filler-driven). This suggests the existence of some sort of gap search procedure driven by the identified potential filler(s). The nature of the cognitive processes at play during gap search is still being explored. More recent theories go further to interrogate the role of maintenance and retrieval mechanisms (Kim et al., 2020) and discourse (Yano & Koizumi, 2021) in filler-driven parsing.

Further evidence has also demonstrated that this procedure varies with the semantic role of the relevant filler phrase. Using a Self Paced Reading Test (SPRT), Crain & Fodor (1985) have demonstrated that reading slows down around the object noun phrase (NP) in WH-questions such as "who", "what" or "when", compared to corresponding declarative sentences.

(1) a. Who had the little girl expected us to

sing those stupid French songs for – at Christmas?

b. The little girl had expected us to sing those stupid French songs for Cheryl at Christmas.

Example (1) taken from Crain & Fodor (1985), as cited in Stowe (1986, p. 229), demonstrates this semantic difference. Their findings show that reading is slower at and around the object NP (in this example, "us") in the WH-condition (sentence (1-a)), as the WH-phrase at the start of the sentence creates an expectation of an object NP gap. The presence of an object NP after the initial WH-phrase before the reader reaches the actual gap (indicated by an en-dash) contradicts that gap hypothesis, and thus the parsing takes longer, creating a so-called parsing slowdown effect. This suggests that the expectation of the gap can influence parsing, depending on the semantic role of the expected filler.

Stowe (1986) then explores this finding further by testing the effects of different phrases of different semantic roles in gap filling of WH-constructions. She replicates previous findings about object NPs, and expands them to gaps in prepositional object roles and subject roles. Her findings show that the processing slowdown effect is not present in subject NPs, suggesting that the WH-phrase is not associated with the subject in the same manner as it is associated with the object. However, they do show the same slowdown effect for prepositional objects as for object NPs, at the gap location created next to the preposition.

This finding is significant because the effect is present even when the preposition is distant from the WH-phrase which starts the filler search, creating a dependency between the proposed filler and the gap which may only be resolved as the preposition is reached and the gap is identified, long after the search for the filler has begun. Such dependencies are called long-distance dependencies, or LDDs. LDDs are formed when the WH-phrase (or other filler phrase) creates an object filler search at the start of the sentence, which is then only resolved by the presence of the preposition indicating the prepositional object gap later in the sentence, the two being separated by several clauses. As the distance between the WH-phrase and the gap can be increased almost indefinitely with recursive embedding, this search can be lengthened significantly, and still the slowdown effect is present.

(2) a. What is Mary hitting –?

b. What is Mary hitting that woman with –?

c. What did Tom think that Mary was hitting that woman with –?

d. What did Harry say that Tom thought that Mary was hitting that woman with –?

Example (2), taken from Stowe (1986, p. 227), shows how LDDs can be created and lengthened almost indefinitely in English in the case of prepositional object gaps (indicated by an en-dash). Sentence (2-a) shows an object gap, but the subsequent sentences use prepositional objects, which will always come after the recursively embedded clauses, to form LDDs in increasingly longer sentences. This suggests that the gap search procedures are not only dependent on semantic and syntactic context, but they are present throughout the parsing of the whole sentence, and identified dependencies are kept track of for significant amounts of time until they are resolved. This paints the picture of a top-down process of sentence parsing which is reactive to context and expectations formed by identified fillers, and active throughout the whole sentence beyond local components.

1.2 Further considerations

While these findings are significant, they are not conclusive, and there is much room for further research on the topic. The functioning of LDDs in languages other than English is still being explored. In order to develop a more universal theory of language processing, other languages with their own unique forms of handling semantic and syntactic context should be taken into account. For example, Aoshima et al. (2004) and Yano & Koizumi (2021) have researched the LDD phenomenon in a head-final language, namely Japanese. Another language of note is Dutch, which was the language examined by Frazier & d'Arcais (1989) in their theory of gap filling. Dutch has WH-constructions that use prepositional objects highly similarly to English, for example with the word "waar", used similarly as the English "that" or "where". However, the findings from Stowe (1986) about WH-phrase-driven

LDDs have not been applied to Dutch yet. Additionally, the Dutch language comes with unique linguistic features that may also form LDDs that cannot exist in English, the status of which is still an open avenue of research.

The pronoun "er" in Dutch, for example, can refer to prepositional objects, creating gaps across long distances, including referring to an object in a different sentence, and is thus a possible candidate for forming similar LDDs. The pronoun has a high versatility and can be used for many functions, including referring back to a previous entity in discourse or introducing new information (used similarly to "daar", the the equivalent of the English "there"). "Er" has been shown to act as a higher-level expectancy monitor (Grondelaers et al., 2009) when introducing new information in a sentence, modulating expectations and easing sentence parsing. And when it acts as a way of referring back to previous information, it can refer to a missing prepositional object. Thus, it may be possible that it can play a role as a prepositional object gap filler as well. This forms a promising avenue for new research on LDDs specific to the Dutch language.

1.2.1 Eye-tracking

Additionally, one other less explored avenue of research in this topic is the variation of methodologies. The previously mentioned "slowdown effect" has been measured using SPRT tests, which comprise of words being read one-by-one at the pace of the reader. Thus, the way to measure whether reading slowed down at or near the target NP was dependent on whole-word differences, with the Region Of Interest (ROI) where the reading speed was measured being defined as "at and immediately following the [target] noun phrase" (Stowe, 1986, p. 229). Different methodological approaches like eye-tracking may allow for a more precise definition of the ROI corresponding to the parsing processes related to the target NP, and better account for other known parsing effects which may affect measurements like the spillover effect (Vasishth, 2006) (see Section 2) by measuring eye fixations directly instead of the time it takes the reader to move on from a given word by pressing a button.

Eye-tracking also offers the advantage of simulating the natural conditions of reading more accurately, and capturing more and richer data about reading, making numerous different eye movement measures available to the researcher to analyse (Wong & Barcroft, 2024, p. 366). Instead of only gathering data on the time it takes the participant to move to the next word, it captures every fixation and saccade of the eye, providing several possible metrics of analysis such as the number of fixations on the ROI, the average duration of fixations on the ROI, and the existence or length of regressions to previous parts of the sentence after the ROI has been reached (Schotter & Dillon, 2025). The latter provides valuable data on backreading and regression behaviour which cannot be obtained using methods like SPRT.

1.3 Research question and hypothesis

As such, I have devised an experiment which aims to expand the findings of Stowe (1986) on WH-phrase LDDs in the English language to the analogous WH-word "waar" in Dutch, as well as exploring another possible candidate for LDD formation unique to Dutch - "er" - using eye-tracking to measure the slowdown effect. This experiment was carried out in parallel to an SPRT study by de Pooter (2025)*.

In this experiment, I intend to investigate the following research question: Do sentences in Dutch constructed with "er" and "waar" to create prepositional object LDDs create a slowdown effect at the gap location ROI (see Section 2.2.2) compared to a corresponding baseline sentence with no gap? If such a slowdown effect exists and shows similar processing speed patterns as other LDDs in English, it would demonstrate an ongoing WH-phrase-driven and "er"-driven filler search process for a prepositional object gap in Dutch, which has not been demonstrated before.

My hypothesis is two-fold: I have a strong hypothesis that there will be a slowdown effect for "waar" sentences compared to the baseline sentences. This hypothesis is deemed as strong, because it has been demonstrated in English for WH-

*Many elements for this project were done in collaboration (see Section 2.2). For clarity and transparency, when discussing elements of the research done collaboratively, I will be using "we", and when I did something alone, I will be using "I" throughout this paper, as well as specifying further detail when necessary.

words by Stowe (1986), which are highly comparable to “waar” both in terms of semantic and syntactic roles in a sentence. The weaker form of the hypothesis, is that there will be a slowdown effect in both “er” and “waar” sentences compared to the baseline sentences. This hypothesis is weaker because there is no comparable pronoun in English and “er” has not been proven to form LDDs thus far. However, due to the syntactic role of “er” in creating gaps for prepositional objects, it is hypothesised that its inclusion will also create an LDD which starts a filler search and results in the same slowdown effect.

2 Methods

2.1 Experiment design

The experiment format followed a sequence of target and filler sentences being presented as stimuli to the participants one after the other, with the eye-tracker tracking their gaze as they read through them. Each participant was assigned a pre-prepared list of target sentences. 12 such lists were prepared based on the 12 sentence sets. The assignment of stimuli to participants followed a Latin square design as seen below.

Participant:	**p0**	**p1**	**p2**	**p3**	**p4**	**p5**
set 1	1b	1w	1e	1b	1e	1w
set 2	2w	2e	2b	2e	2w	2b
set 3	3e	3b	3w	3w	3b	3e
set 4	4b	4w	4e	4b	4e	4w
set 5	5w	5e	5b	5e	5w	5b
set 6	6e	6b	6w	6w	6b	6e

Table 2.1: Example of Latin square distribution of 6 stimuli sets for 6 participants. “b”, “w” and “e” refer to “baseline”, “waar” and “er” conditions respectively.

The variable measured was the duration the participant spent on the Region of Interest (ROI) associated with the gap location (for details, see Sections 2.2.2 and 2.3.3) before moving on from it and reading further, also known as go-past time (Schotter & Dillon, 2025). The goal behind this measurement was to ascertain how long the brain spends processing the prepositional gap, as the time spent fixating on the ROI is representative of processing time (Just & Carpenter, 1980). This variable was compared between subjects and between sentence sets.

2.2 Materials

24 sentence sets were made collaboratively, as this experiment was performed in parallel to a self-paced reading experiment by de Pooter (2025), using parallel stimuli constructed together. The sentence sets each comprised of a “baseline”, an “er” version and a “waar” version (Details given in Section 2.2.1). Of those, half were assigned to my experiment, and half to the parallel self-paced reading experiment. As such, 12 target stimuli were shown to each participant along with filler stimuli inserted into the Latin square design pseudorandomly. 36 filler stimuli were also made collaboratively, with half of them being assigned to my experiment. As such, the total number of stimuli for each participant was 30 (12 target + 18 filler). The process of pseudorandomisation selected filler stimuli to insert in a random order into the Latin square columns, but ensured that no more than 2 target stimuli were ever shown consecutively, and no more than 2 filler stimuli were ever shown consecutively.

2.2.1 Target stimuli

Each sentence set consisted of one sentence for each condition. The sentence structures were standardised to ensure the least possible lexical and grammatical variability within sets aside from the rephrasing necessary to achieve the relevant condition. All baseline sentences were in present perfect tense using “hebben”, “zijn” or modal verbs as auxiliary. Additionally, in order for the sentences to include a prepositional object, the verbs used were chosen from a list of Dutch verbs with fixed prepositions (nt2.nl, 2017). The template for the structure of baseline sentences can be seen below.

Example baseline sentence:
“De studenten hebben vorige week veel geld aan dure kostuums voor Halloween besteed in de feestwinkel.”
“The students spent a lot of money on expensive

Halloween costumes at the party store last week."

Sentence part	Structural role
De studenten	subject
hebben	auxiliary
vorige week	adv. clause (1 or more)
veel geld	object (optional)
aan	preposition
dure kostuums	prepositional object
voor Halloween	adv.clause (optional)
besteed	lexical verb
in de feestwinkel.	adv.clause

Table 2.2: Example and structure of baseline condition sentences

The structure of the sentence outlined by Table 2.2 was the same for every baseline sentence. At least one adverbial clause was always present after the auxiliary. Some sentences had optional objects as well as prepositional objects, always in the same position before the preposition. Some sentences had an optional adverbial clause after the prepositional object, added as needed to ensure the length to create the LDD effect. All sentences ended with an adverbial clause after the lexical verb, to account for the wrap-up effect, (Just & Carpenter, 1980, p. 345; Schotter & Dillon, 2025, p. 5) which states that reading speed slows down at the end of a sentence. This ensures that the slowing down of reading speed due to wrap-up procedures does not fall into the processing ROI associated with the gap and interact with the slowdown effect to be measured.

For "waar" sentences, in order to ensure the most similarity possible to the baseline sentence, we decided not to phrase the WH-construction in the form of a question, but rather a declarative. This required the construction of a relative WH-clause to create the LDD between the WH-phrase and the prepositional object gap embedded within the clause, moving the previous prepositional object out to be the subject of the matrix sentence at the beginning. The structure of the relative clause corresponded to the structure of the respective baseline sentence, the main difference being the absence of the prepositional object which creates the gap, as seen in Table 2.3.

Example "waar" sentence:

"De Halloweenkostuums, waar de studenten vorige week veel geld aan hebben besteed in de feestwinkel, waren duur."
"The Halloween costumes that the students spent a lot of money on at the party store last week were expensive."

Sentence part	Structural role
De Halloweenkostuums	matrix subject
waar	WH-phrase ("waar")
de studenten	subject
vorige week	adv.clause (1 or more)
veel geld	object (optional)
aan	preposition
–	prep. object gap
hebben besteed	verb+auxiliary
in de feestwinkel	adv.clause
waren duur.	matrix verb phrase

Table 2.3: Example and structure of "waar" condition sentences

For "er" sentences, we similarly required the prepositional object to move outside of the relevant clause to create a gap. We addressed this by making a separate context sentence which introduces the prepositional object, and then the target sentence which refers back to it using "er", and includes the gap. Similarly to the previous condition, we ensured the structure of the target sentence corresponded to the baseline sentence as much as possible, as shown below.

Example "er" sentence with preceding context: "Iedereen heeft voor het feestje een duur Halloweenkostuum gekocht. **De studenten hebben er vorige week veel geld aan besteed in de feestwinkel.**"
"Everyone bought an expensive Halloween costume for the party. The students spent a lot of money on it last week at the party store."

Target sentence part	Structural role
De studenten	subject
hebben	auxiliary
er	er (refers to prep. obj)
vorige week	adv.clause (1 or more)
veel geld	object (optional)
aan	preposition
–	prep. object gap
besteed	lexical verb
in de feestwinkel	adv.clause

Table 2.4: Example and structure of "er" condition sentences

All prepared target stimuli sets followed this format to ensure minimal structural variability. In the process of creating new sentences we ensured a use of diverse verbs, subjects and adverbial phrases. In order not to interfere with the filler search process, no prepositions were included in any of the adverbial phrases occurring before the gap location. The draft set of sentences were reviewed by a native speaker to ensure readability and authenticity.

2.2.2 ROI specification

For each target stimulus, I determined the location where the lexical processing procedures associated with the prepositional object (or its gap) would occur during reading. This location is called the Region of Interest (ROI). For baseline sentences, there is no gap, so the relevant ROI is the prepositional object. For "waar" and "er" sentences, it is the prepositional object gap. In all three cases, the ROI comes immediately after the associated preposition, either followed by a gap or its object. As the region I am interested in is not a word but a gap left behind, there is no exact measure for the number of characters this encompasses. Thus, I defined the area around the character following the end of the preposition to be the ROI. I defined "around" to be 5 characters before or after that character as that encompasses the maximum area around which a fixation can be deemed as a fixation on the gap location and not a neighbouring word in my estimation.

However, I must also account for known cognitive effects on the interaction between reading and processing, such as the spillover effect. Post-lexical integration processes like the gap filling procedure are known to lag behind eye movements used for reading (Reichle, Liversedge, et al., 2009; Reichle, Warren, & McConnell, 2009 as cited in Schotter & Dillon, 2025, p. 6), and as such the eye will move on to the next part of the sentence after the gap location before processing related to the gap location has finished. Therefore, I decided to extend the window of the ROI by 5 characters after the gap, in order to accurately capture the window where processing would occur.

As such for all sentences my ROI is defined as a window of character indices 15 characters long, comprising of the 5 characters before the character following the preposition, and 10 characters after, as shown in Figure 2.1.

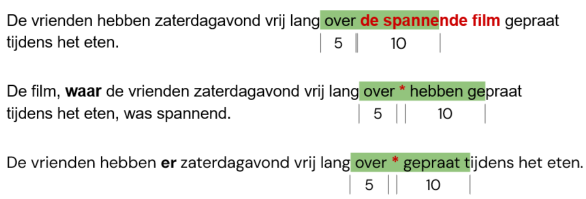

Figure 2.1: Visualisation of the number of characters in the ROI on an example set of sentences. The prepositional object or gap is shown in red, and the ROI window is highlighted in green.

The goal behind this choice is to be precise enough to capture only the region where the gap (rather than neighbouring words) is being processed, but broad enough to account for the fact that fixations which process the gap will not always occur exactly at the gap location between two words, as eye movements are prone to a degree of randomness from over- and undershooting their target (Conklin et al., 2023, p. 77).

2.2.3 Filler stimuli

As the target stimuli followed a distinct pattern across each condition, we also included extraneous filler stimuli to ensure that pattern would not be recognised, and the true focus of the experiment could not be deduced by its participants. We created three types of filler stimuli. We employed the use of OpenAI's GPT-3.5 Turbo model (OpenAI, 2025) via ChatGPT in the creation of these sentences, which we then selected and modified for authenticity, diversity, and similar length to target sentences.

The first type of filler sentences were sentences with major syntactical errors. To achieve this, example sentences in Hungarian provided by ChatGPT were translated word-for-word into Dutch. As Hungarian has radically differing syntax rules compared to Dutch, this resulted in sentences where the intended meaning could be devised, but was obviously syntactically incorrect.

The second type of filler sentences were sentences with agreement errors. These were otherwise grammatically correct example sentences in Dutch generated by ChatGPT which were modified to contain gender disagreement and number disagreement between subjects and verbs or nouns and their adjectives, and incorrect pronouns.

The third type of filler sentences were grammatically correct Dutch example sentences. Some were sourced from a Dutch book (Veer, 2007), some were generated by ChatGPT. For some, context sentences were also prepared, to resemble the context + main sentence format of "er" condition stimuli.

2.2.4 Eye-tracker

An Eyelink Portable Duo™ was used for eye-tracking. A screen illumination of 100% was used as well as the head-stabilised setting. 13 calibration dots (the maximum setting) were used during calibration. Advanced settings were all left on default. The eye-tracker tracked the left eye for all participants but one, for whom calibration could not obtain a "GOOD" rating tracking the left eye. Instead, right eye tracking was used for that participant.

2.3 Procedure

Participants were seated in front of the eye-tracker and the display monitor, and had their head stabilised with a head brace to ensure they did not move their head during the experiment. Participants with glasses were asked to clean their lenses beforehand. A calibration and validation of the eye-tracker was performed at the start of the experiment for each participant, and the experiment proceeded once a calibration rating of "GOOD" was achieved. After that, participants were instructed to read the sentences on the monitor and press the space bar once they had completed reading it. In the case of "er" sentences, a context sentence would be shown above the main sentence, which they were instructed to read beforehand. A fixation dot would then be displayed in the centre of the screen for 2 seconds, which participants were instructed to return their gaze to between trials. The screen would then display the next sentence, and the process would loop until all stimuli were shown.

This procedure was explained in written instructions, and a practice round with five trials showing example sentences was conducted to ensure all participants understood and were used to the experiment procedure before it started.

2.3.1 Participants

15 participants were selected, the majority being students studying at the University of Groningen. All participants were native speakers of Dutch with no known conditions affecting reading and processing speed (i.e. dyslexia) and normal or corrected vision. No further demographic data was collected as the analysis method already corrected for any variation between subjects (see section 3.2). All participants signed a consent form ahead of the experiment, which outlined the experiment procedure and stated that the experiment posed no known risks.

2.3.2 Experiment software and specifications

The experiment ran in a full-screen OpenSesame (Mathôt et al., 2012) environment using the xpyriment back-end and the OpenSesame eye tracker modules. Text was displayed on a 1920x1080 pixel monitor with white text on a black background, in a monospace, 24 point font. A Python script was used to run the experiment, which displayed the prepared sentences for each participant in the centre of the screen (in the case of "er" sentences, with the context sentence displayed 120 pixels above), and sent relevant variables about the current trial to the eye tracker. Use of the Sigmund AI research assistant, a tool to aid OpenSesame development (Mathôt, 2025), was employed over the course of the development of the experiment environment.

2.3.3 Data post-processing

Another Python script was used to process eye-tracker output. Variables logged during each trial were used to match fixation data to the corresponding sentences. The start and end character indices of the ROI in the given sentence as well as font size and character width information was then used to calculate the pixel positions of the ROI on the screen, and determine whether fixations were within the ROI. Some fixations within the ROI were deemed invalid and filtered out, because they occurred at the immediate beginning of the trial when reading had not started yet (there were no fixations at the start of the sentence previously). Such erroneous fixations were likely a result of the gaze being centred on the screen between trials, and not part of any lexical processing.

Another calculation was then made to determine ROI go-past time. This was done by extracting the starting time of the first fixation within the ROI and subtracting it from the starting time of the first subsequent fixation to the right of the ROI. Go-past time was then compared as a representation of processing time across conditions to analyse results.

3 Results

3.1 Data analysis

In order to gauge patterns in the data, the fixation dataset obtained after post-processing was analysed and go-past time was compared across conditions. Out of 180 target trials (15 participants x 12 trials), 175 valid trials were analysed as data points. 3 trials did not have any fixations within the ROI, and were thus deemed invalid for the purposes of analysis. Go-past time was impossible to calculate for 2 further trials, as there were no fixations to the right of the ROI after any within-ROI fixations. The go-past times were grouped both by participant and by set, to reveal possible patterns in the data. A visualisation can be seen in Figures 3.1 and 3.2.

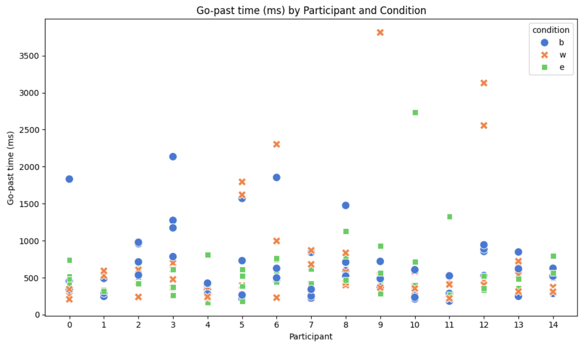

Figure 3.1: Go-past time grouped by Participant and Condition

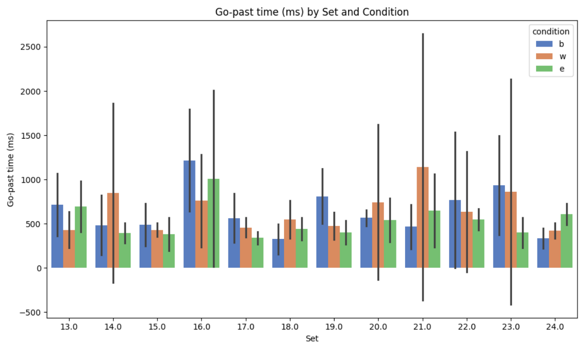

Figure 3.2: Go-past time grouped by Set and Condition

Visual analysis of Figures 3.1 and 3.2 does not reveal any discernible patterns in the data across conditions. The data shows no notable features aside from individual differences and outliers. For example, the highest go-past time was in a "waar" condition trial by participant 9. It lasted over 3500ms while the majority of data points show a go-past time of under 1500ms. After some additional data analysis into the order and position of fixations, this was revealed to be due to the participant rereading the start of the sentence multiple times before moving on from the ROI. Such regressions are expected and are intentionally included in the calculation of go-past time as they reflect the time needed to resolve processing difficulty (Clifton Jr et al., 2007 as cited in Schotter & Dillon, 2025, p. 6). Similarly, when the data was grouped by set, there was high variation across all conditions, and no particular condition stood out in terms of having a significant effect on go-past time. The highest average go-past time was for set 16 in the "baseline" condition and set 21 in the "waar" condition, but as shown by the large error bars this is likely due to individual outlier data points with high go-past time, where the participant regressed for a long time during reading. Visual inspection of the data revealed no conclusive insights into the effect of condition on go-past time.

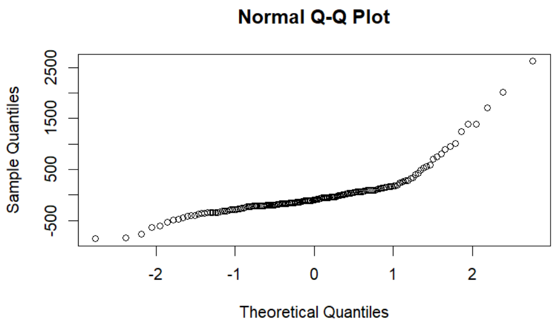

Figure 3.3: Normal Q-Q plot of the residuals of the model before the log transformation showing heteroskedasticity

3.2 Statistical modelling

In order to represent the effect of each condition on go-past time, a linear mixed effects model was constructed using the lme4 package in R (R Core Team, 2024; Bates et al., 2015), following the steps outlined by Winter (2013) to fit all 175 data points. The model query can be seen below:

$$experiment\_model = log(go\_past\_time) \sim condition + (1 + condition|participant) + (1 + condition|set)$$

Go-past time was predicted based on condition, with random intercepts and random slopes by participant number and set number. This accounts for variation between participants (as differing baseline reading speeds are expected) and sentence sets (which use words of varying frequencies and may include unexpected confounding factors in sentence structure) as well as the effect of the conditions on these variables.

A natural logarithmic transformation was applied on go-past time in order to significantly reduce heteroskedasticity present in the residuals, as can be seen in Figures 3.3 and 3.4. The final model with the log transformation still exhibits some slight signs of heteroskedasticity at the tail of the residuals, but to a negligible extent which still accurately captures the variation within the data.

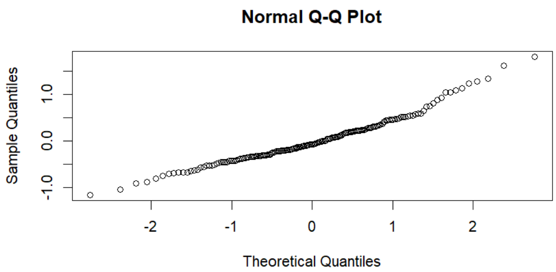

Figure 3.4: Normal Q-Q plot of the residuals of the model after log transformation showing only slight heteroskedasticity at the tail

A summary of the model showed only a slight, non-significant effect of both the "er" and "waar" conditions on go-past time.

```
Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.2317 -0.6282 -0.1512  0.4816  3.4575

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 participant (Intercept) 0.070134 0.26483
             conditione  0.050172 0.22399  -1.00
             conditionw  0.036751 0.19171   0.00  0.00
 set         (Intercept) 0.033266 0.18239
             conditione  0.009867 0.09933  -0.86
             conditionw  0.008643 0.09297  -0.99  0.93
 Residual                0.273577 0.52305
Number of obs: 175, groups:  participant, 15; set, 12

Fixed effects:
            Estimate Std. Error t value
(Intercept)   6.2668     0.1100  56.976
conditione   -0.1063     0.1167  -0.911
conditionw   -0.0946     0.1116  -0.848

Correlation of Fixed Effects:
           (Intr) condtn
conditione -0.772
conditionw -0.493  0.412
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Figure 3.5: Summary output of the Linear Mixed effects model, showing the effect of conditions on go-past time

As seen in Figure 3.5, the "er" condition lowered the log(go-past time) by 0.0163 ms ± 0.1167 compared to the baseline condition (the intercept) with a t value of -0.911, and the "waar" condition lowered log(go-past time) by 0.0946 ms ± 0.1116 compared to the baseline condition with a t value of -0.848. Neither of the t-values prove significant within the model itself. However, a further comparison with a null model with the effect of the condition removed is needed to ascertain the true significance of the presence of any non-baseline condition on go-past time.

3.3 Significance tests

In order to test whether the effect of non-baseline conditions was significant on go-past time, a likelihood ratio test of two models was carried out using the anova() function in the base package in R (R Core Team, 2024), comparing the linear model to a null model with the effect of the condition removed (as outlined by Winter (2013)). The query of the null model can be seen below, removing only the condition:

$$experiment\_model = log(go\_past\_time) \sim (1 + condition|participant) + (1 + condition|set)$$

```
> anova(experiment.model.null, experiment.model.c)
Data: base_dataset
Models:
experiment.model.null: log(time_to_right) ~ (1 + condition | participant) + (1 + condition
| set)
experiment.model.c: log(time_to_right) ~ condition + (1 + condition | participant) + (1 + c
ondition | set)
                      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
experiment.model.null   14 324.68 368.98 -148.34   296.68
experiment.model.c      16 327.55 378.19 -147.78   295.55 1.1221  2     0.5706
```

Figure 3.6: Results of the likelihood ratio test between the null model and the full model including the condition

As shown by Figure 3.6, results showed that the variation of conditions did not significantly affect go-past time ($\chi^2(2)$=1.1221, p=0.5706).

This result is not significant, and therefore does not disprove the null hypothesis that "er" and "waar" conditions do not have any effect on go-past time at the gap location.

4 Conclusion

In this study, I explored the relationship between reading speed and prepositional object gaps formed by LDDs in Dutch. Specifically, I measured the time it takes for readers' fixations to go past the ROI associated with a prepositional object or prepositional object gap location in three different conditions: a baseline sentence with a prepositional object, a "waar" sentence with the gap embedded in a relative clause, and an "er" sentence where the object is established as context and a gap is created when referring back to it. I then created a statistical model of the go-past time to ascertain the effect of both the "er" and "waar" cases compared to the baseline case. The experiment did not find a significant slowdown effect in either condition, and has obtained a null result. Neither the strong version of the hypothesis, that the "waar" condition would cause a reading slowdown effect, nor the weak version, that both the "er" and "waar" conditions would cause a slowdown effect, could be proven. There are two possible avenues to explore for the reason behind this null result. Either the anticipated effect is not present, in which case the hypothesis was flawed, or the effect does exist, but the experiment was not sufficient in its design and as a result the effect could not be found.

4.1 Discussion of hypothesis

My hypothesis stated that based on evidence from previous literature (Stowe, 1986; Crain & Fodor, 1985; Frazier & d'Arcais, 1989), sentences constructed to form an LDD between proposed fillers and a prepositional object gap will lead to a slowdown effect at the gap location as it induces a filler search process during which the filler creates an expectation of a prepositional object, especially in the case of the filler being a WH-phrase. Since the slowdown effect could not be evidenced, this suggests that no filler search procedure took place. It may be the case that "er" and "waar" do not act as viable filler phrases in creating LDDs for prepositional object gaps in the same way that WH-phrases in English do, and they do not trigger the same underlying filler search processes that were posited by Stowe (1986), as the linguistic mechanisms involved are not equivalent. If the hypothesis is indeed flawed, the alternative explanations of how the gap is accounted for may be different in the cases of the "er" and "waar" sentences.

In the case of "waar", a flaw in the hypothesis is less likely to be true, since the word is used as a relative pronoun in the same way that WH-words such as "what" or "where" are used in in English. The structure of relative clauses with WH-words in Dutch is highly analogous to sentence structures in English, where, as previously mentioned, a slowdown effect has been found. As such, I will look to methodological explanations for the null result instead.

In the case of "er", the equivalence to WH-phrase filler search processes is not so apparent, and as such more research is needed to explore what cognitive processes take place in linking it to a prepositional object, or any other constituent it refers to among its several use cases. The "er" pronoun can be used to introduce new information, or, alternatively, to constrain the expectations on, or refer back to, existing information. In some cases, its use is analogous to the English use of the word "there" when used as a function word to introduce new clauses, and in others it acts as a personal pronoun like the English "it". In this experiment, it was used as a personal pronoun referring to the previously established object, and creating the expectation of a prepositional object in the sentence. This is less obviously analogous to previous methods of exploring a filler search process, and may involve different underlying parsing processes than expected. One other possible effect of "er" that has been explored is its role an expectancy monitor, which can, in some cases, speed up reading instead of slowing it down (Grondelaers et al., 2009). Future hypotheses about the role of "er" in discourse should account for and question such effects.

Overall, a methodological explanation for the null result is more likely for the case or "waar", while in the case of "er" the hypothesis itself is more uncertain. More research into the nature of the pronoun "er" is also necessary, in order to understand its effects on processing and expectations during reading in different use cases.

4.2 Methodological limitations and future research directions

The other possibility for the null result is that the experiment design was flawed and could not find the effect present. I will explore possible avenues for such flaws and ways they may be addressed in the future.

4.2.1 Statistical power

Due to time constraints, no power analysis was conducted before the experiment, and as such the statistical power required to find the slowdown effect is unknown. The 15 participants and 175 trial data points among them may not have been enough to find a statistically significant result. A post-hoc power analysis could still offer this insight, and is a possible avenue for a future revisiting of this study.

4.2.2 Experimental design of the ROI

Some experimental design considerations were based on reasonable assumptions about the most effective approach. Some of these assumptions may be flawed and as such there is room for exploration in the approach to experimental design. For example, the choice of location and window width for the ROI (see: Section 2.2.2) could be defined differently, including a wider or narrower window before or after the prepositional object gap, based on different assumptions about the precision of the gap location and ways of accounting for different cognitive factors like the spillover effect. Current eye-tracker data could be re-analysed with different ROI windows to obtain deeper insight into the data.

Additionally, data analysis on different metrics of measurement rather than go-past time determined from the start of within- and subsequent out-of-ROI fixations could also be insightful. An advantage of the eye-tracking methodology is that the data obtained for this experiment is rich and open to many methods of analysis which may be revisited in the future. Saccade data, rather than fixation data, for example, was not used at any stage of analysis, but may offer some insight into reading behaviour during possible filler search processes.

4.2.3 Design of stimuli

The construction of target stimuli may also be improved. Gathering new data using different stimuli and refined stimulus design considerations may result in more conclusive results. This may include an entirely new sentence structure, or more minor changes designed to aid the creation of a filler search process. For example, including "waar"-question phrases instead of declaratives with a relative clause in the stimuli set could be a more organic way of producing the slow-down effect, and would more closely mirror the experiment by Stowe (1986) which has found evidence for a filler search process. An example of such possible question phrases can be seen below:

"Waar hebben de vrienden zaterdagavond vrij lang over — gepraat tijdens het eten?"
"What did the friends talk quite a while about - on Saturday evening during dinner?"

4.2.4 Experimental factors

One possible confounding effect during the experiment was the fact that one participant's data (and as such, 12 data points) was recorded by tracking the right eye instead of the left eye unlike all other participants, due to calibration difficulties. This goes against the recommended practice of tracking either the same eye for all participants or performing an eye dominance test and tracking participants' dominant eye in monocular tracking to ensure standardisation (Conklin et al., 2023, p. 59). While this may have introduced some unexpected variation into the data in the case of one eye being more dominant over the other, due to the low significance scores obtained it is unlikely that it has altered the overall result.

References

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, *51*(1), 23–54.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Clifton Jr, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. *Eye Movements*, 341–371.

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2023). Eye Tracking: A Guide for Applied Linguistics Research. *Southern African Linguistics and Applied Language Studies*, *41*(3), 385–388.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers. *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*, 94–128.

de Pooter, G. (2025). Exploring Long-Distance Dependencies in Dutch: The Role of 'Waar' and 'Er' [Bachelor's Thesis, Artificial Intelligence]. Retrieved from `https://fse.studenttheses.ub.rug.nl/id/eprint/34652`

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9*(3), 427–473.

Fodor, J. D. (1979). Superstrategy. *Sentence Processing*, 250–279.

Frazier, L., & d'Arcais, G. B. F. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*(3), 331–344.

Grondelaers, S., Speelman, D., Drieghe, D., Brysbaert, M., & Geeraerts, D. (2009). Introducing a new entity into discourse: Comprehension and production evidence for the status of Dutch er "there" as a higher-level expectancy monitor. *Acta Psychologica*, *130*(2), 153–160.

Jackendoff, R. S., & Culicover, P. (1971). A reconsideration of dative movements. *Foundations of Language*, 397–412.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, *87*(4), 329.

Kim, N., Brehm, L., Sturt, P., & Yoshida, M. (2020). How long can you hold the filler: Maintenance and retrieval. *Language, Cognition and Neuroscience*, *35*(1), 17–42.

Mathôt, S. (2025). *Sigmund AI.* Retrieved 2025-03-19, from https://sigmundai.eu/about

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314–324.

nt2.nl. (2017). *Lijst werkwoorden met vast voorzetsel.* Retrieved 2025-04-03, from https://www.nt2.nl/documenten/in_zicht/lijst_werkwoorden_met_een_vast_voorzetsel.pdf

OpenAI. (2025). *GPT-3.5 Turbo Documentation.* Retrieved 2025-03-19, from https://platform.openai.com/docs/models/gpt-3.5-turbo

R Core Team. (2024). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences*, *13*(3), 115–119.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using EZ Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*, 1–21.

Schotter, E. R., & Dillon, B. (2025). A beginner's guide to eye tracking for psycholinguistic studies of reading. *Behavior Research Methods*, *57*(2), 68.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245.

Vasishth, S. (2006). On the proper treatment of spillover in real-time reading studies: Consequences for psycholinguistic theories. In *Proceedings of the International Conference on Linguistic Evidence* (pp. 96–100).

Veer, W. (2007). *Wat vliegt daar? Vogels rond het huis.* Z.n. Paramaribo. (accessed via DBNL (KB, national library))

Wanner, H., & Maratsos, M. P. (1978). An ATN approach to comprehension. In *Linguistic Theory and Psychological Reality.* M.I.T. Press.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint arXiv:1308.5499*.

Wong, W., & Barcroft, J. (2024). *The Routledge Handbook of Second Language Acquisition and Input Processing.* Taylor & Francis.

Yano, M., & Koizumi, M. (2021). The role of discourse in long-distance dependency formation. *Language, Cognition and Neuroscience*, *36*(6), 711–729.