university of groningen / faculty of science and engineering

# The effects of prepositional 'er' on reading speed in Dutch, an eye tracking experiment.

Bachelor's Project Thesis

Jurjen van der Weit, s3658619, j.e.van.der.weit@student.rug.nl,
Supervisor: Dr S.M. Jones

**Abstract:** This paper studies the effect of the prepositional function of the Dutch word 'er' on reading speed. This was done by conducting a Dutch self paced reading experiment using eye tracking. The effect was tested by showing participants sentences with prepositional 'er' and a baseline version where 'er' was replaced by the subject it referred to and comparing the reading times of the preposition including a spillover region. From previous research we expect that prepositional 'er' will behave much like a Long Distance Dependency and therefore increase the time needed to read the preposition. The findings of this paper show that there is indeed a significant slowdown in reading speed at the preposition when the prepositional 'er' is present in the sentence versus when it is not.

## 1 Introduction

In Dutch, the word 'er' has many different meanings. While in many cases it can be translated as 'there', sometimes it cannot be translated at all. Er shows four different functions as categorized by Odijk (1993): Existential $er_X$, Locative $er_L$, Prepositional $er_P$ and Quantative $er_Q$. There are multiple previous researches on the distributions of the use of 'er' and its syntactic constraints, Bech (1952), Bennis (1986), Van Riemsdijk (1978), Odijk (1993), Neeleman and Van de Koot (2017), Donaldson (2008), Grondelaers, Speelman, Drieghe, Brysbaert, and Geeraerts (2009), Klooster (2014), Webelhuth and Bonami (2019), from which we know that 'er' generally only occurs once and twice at a maximum in a sentence, but that it can fulfill more than one function at the same time. Grondelaers et al. (2009) and Grondelaers (2020) show that for certain functions of 'er', the reading speed is increased. While this study focuses only on the prepositional function of 'er', all four functions of 'er' will be explained to better understand the context of this research.

### 1.1 Existential $er_X$

Existential $er_X$, also called presentative $er_X$, can be found at the beginning of a clause where the subject is indefinite (1) or as the subject of an impersonal passive sentence (2). $er_X$ must also appear in the middle of the sentence when the start is occupied by a non-subject constituent and there is no other explicit subject but is optional when there is an explicit subject (3) Jones (2020).

(1) Er ging een man naar de bakker.
There went a man to the bakery.
'A man went to the bakery.'

(2) Er wordt gegeten.
There is becoming eating.
'There is eating.'

(3) Waar wordt er aan het spoor gewerkt?
Where is becoming there on the track worked?
'Where will there be maintenance on the tracks?'

### 1.2 Locative $er_L$

Locative $er_L$ replaces a locative phrase in a sentence and can be replaced by *daar/hier* "there/here". $er_L$ cannot be at the start of a sentence when it is the only function of 'er'. See (4) for an example of how $er_L$ is used.

(4) Ik ben er nooit geweest.
I have there never been.
'I have never been there.'

### 1.3 Prepositional $er_P$

Prepositional $er_P$, also called pronominal $er_P$ can be replaced by *daar/hier* "there/here". $er_P$ is used when the sentence requires a non-human prepositional object and is always combined with a preposition like ***er****naast* 'next to **it**' or ***er****onder* 'under **it**'. $er_P$ is most of the time split from its preposition (5), but is written together when the preposition directly follows 'er' (6).

(5) Je zat er net een beetje naast.
You sat there just a bit next.
'You were off by just a bit.'

(6) Ik kijk ernaar.
I look it towards.
'I am watching it.'

### 1.4 Quantative $er_Q$

Quantative $er_Q$, also known as paritive $er_Q$, is used with a quantity (7). It cannot be replaced by *daar/hier* and only can appear in the middle of a sentence.

(7) Ik heb er drie gekocht.
I have there three bought.
'I have bought three of them.'

### 1.5 Long Distance Dependencies

To better understand why $er_P$ is interesting, Long Distance Dependencies (LDDs) should be explained first. A Long Distance Dependency is a phenomenon where two words are related and both are needed to interpret the meaning of them, but they are not adjacent to each other. In English, this is common for WH-questions (what, where when, etc), where, for example in (8), 'which' and 'buy' have to be interpreted together to understand the sentence.

(8) Which apple did you buy?

In this case, the full interpretation of 'which' is left open until the second part of the LDD is finally parsed and the whole grammatical structure can be understood. The length of the LDD can vary based on the number of clauses and words used. Gibson and Warren (2004) argues that it is costly to keep a LDD open, increasing the time needed to process it when the length of the LDD increases. Phillips, Kazanina, and Abada (2005) argue, however, that shorter LDDs are preferred over longer LDDs and that a reader will try to resolve the LDD at the earliest opportunity. They also argue, just as Gibson and Warren do, that the processing cost of a LDD is distance sensitive and increases if more words have to be processed in between the LDD. Furthermore, it is known from Stowe (1986) that LDDs in general show a slowdown in reading speed when the reader has to parse the second part of the LDD.
While most research about LDDs is focused on English, we know as well that LDDs exist in, amongst others, Icelandic, English and Japanese (Kaplan and Zaenen, 1989) but also in Dutch, Frazier (1987) and (Bouma, 2017).

### 1.6 Research Question & Hypothesis

While all of the four functions of 'er' are interesting and can be used in conjunction with one another, this research project will focus on the prepositional $er_P$ and what its effects are on reading speed. As described before, LDDs decrease reading speed and there seems to be syntactical Long Distance Dependency between $er_P$ and its preposition. Investigating this property of $er_P$ could give some insight into the role of 'er' in a sentence. While there is a small amount of research on the use of $er_P$, Berends, Hulk, and Sleeman (2017), there is little to no research that goes into the effect of $er_P$ on reading speed. The research question is therefore as follows:
"What is the effect of prepositional 'er' on reading speed in a Dutch reading experiment?".
We hypothesize that $er_P$ will take longer to process in comparison to a baseline version of the same sentence without the $er_P$ function. $er_P$ will slow down the reading speeds since participants will either have to think longer about the preposition they are reading or go back in the sentence to reread $er_P$ and make the sentence a cohesive whole. This happens because of the property of LDDs where

$er_P$ will not be resolved until the preposition is read. However, this is based on the results of Stowe (1986) where resolving the LDD slowed down the reading speed. Stowe, however, conducted the experiment in English. We expect that this effect will also be present in Dutch based on the fact that LDDs are present in the Dutch language as found by Frazier (1987) and Bouma (2017). We expect that the results of this study will deviate from the findings of Grondelaers et al. (2009) on the effect of 'er' on reading speed, where 'er' increased the reading speed of sentences. However, Grondelaers et al. focused on the existential $er_X$ and therefore we hypothesize that the different functions of 'er' have different effects on reading speed.
The experiment will make use of eye tracking since this will give us the most accurate results in comparison to a Self Paced Reading experiment (SPR). This is because in eye tracking, the whole sentence can be shown at once, while in SPR chunks of the sentence are shown at a time, avoiding issues with words depending on information from other chunks.

# 2 Methods

To test the hypothesis that $er_P$ will take longer to process than the same sentence without $er_P$, an eye tracking experiment was conducted.

In this experiment participants performed a self-paced reading task on 40 Dutch sentences shown 1 by 1. While reading, a eye tracking camera kept track of the right eye of the participant to be able to record the eye movement data.

## 2.1 Experimental Setting

The experiment took place in the Bernoulliborgh building of the Rijksuniversiteit Groningen. As shown in figure 2.1, participants were seated in front of a computer screen with a 1080 by 1920 resolution and had to place their head in the headrest located at the front of the desk. The brightness of the screen was turned down to 50% to increase the comfort of reading the screen. A camera was placed in their field of vision with the specifications described in section 2.2. The interior of the room was minimal: a white wall as a background and no decoration. A filing cabinet separated the participant and surveyor so no visual contact could be made during the experiment. A keyboard was used to let the participant give feedback whenever they were done reading a sentence and speakers were used during the calibration of the eye tracker to give auditory cues.

**Figure 2.1: Picture of experimental setup**

## 2.2 Camera

The camera used in this experiment is the Eyelink Portable Duo Eye Tracker set to a sampling rate of 500 Hz. The camera was connected to an external laptop that was on the other side of the file drawer for remote control. Before the experiment began, the right pupil of the participant had to be calibrated to the camera for it to know where the participant was looking. This was done using a 9 point calibration grid with validation done right afterwards.

## 2.3 Experiment

The experiment was made in OpenSesame, version 3.3.11 and available **here**. In the experiment, participants had to first have their eyes calibrated to the camera as explained in section 2.2. Afterwards, they were shown a screen with instructions about the amount of time that the experiment would take, how many sentences were going to be shown, when to press the space bar to continue and that they had to look at a drift correction dot in between sentences. In total, 40 sentences were shown to the participant one by one, with a drift correction dot in between that prevented large discrepancies between the recorded data and where the participant was actually looking. All sentences were vertically centered and started at X = 410 on a scale from

0 to 1920. The text was in the mono font, size 18. Whenever a participant was done reading a sentence, they pressed the space bar to continue. At the end a screen was added that indicated that they were finished with the experiment. The whole experiment took about 15 minutes.

## 2.4 Participants

26 participants from the University of Groningen and the Hanze University participated in this study, of which 16 male and 10 female. The age group was 18-24 with a mean age of 21.5. All participants were native Dutch speakers with normal or corrected to normal vision and had no dyslexia. Each participant received 5 euros for participating.

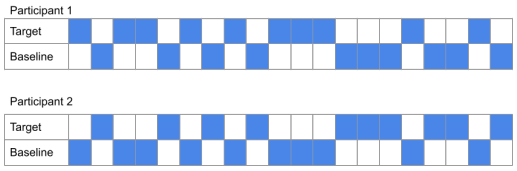


**Figure 2.2: An example of how the sentences for a pair of participants were created. The blue squares show which sentence was shown to the participant.**

## 2.5 Sentences

Each participant was shown 40 sentences drawn from a pool of 80 sentences. These were subdivided in 3 classes. 10 out of 20 target and 10 out of 20 baseline sentences and 20 out of 40 filler sentences were picked using a Latin square design. Each target sentence has a corresponding baseline sentence that is the same but without $er_P$, see section 2.5.1 and 2.5.2.
The data was created in pairs of participants. Within each pair, for the first participant, the 10 target sentences were picked randomly. The baseline sentences then were the left over sentences that were not a corresponding version of the target sentences. For the second participant, the target and baseline sentences were inverted. See figure 2.2 for a graphical interpretation of this process. This process was repeated for every pair of participants to create the whole dataset. Even though the input data was linked, the participants were not linked in any way. A Python script was made to create the structure as explained above, which can be found **here**. This script created separate csv files for each participant and a csv file that kept track of which participant got which sentence. The sentences were not checked for real world plausibility, only grammatical plausibility. The full list of sentences can be found in the appendix.

### 2.5.1 Target

This class of sentences was the main aim of the experiment and contained the $er_P$ grammatical construction. All sentences were constructed using a certain layout. As in (5), 'er' was in the middle of the sentence and only carried out the function of $er_P$. To test the LDD effect of $er_P$, 'er' and its corresponding preposition were spaced 3 to 5 words apart. This gap was kept relatively constant to measure the general effect of LDDs. Since $er_P$ is linked to a certain preposition, no prepositions are used in the gap between $er_P$ and its preposition to prevent the effect found by Stowe (1986), where a slowdown was found at each potential word that could resolve the LDD, in this case a preposition. After the preposition, a chunk of three words minimum is added to account for the spillover effect. This is the effect that there is a delay between reading and processing a word, which will mean an effect of a certain word will be visible later than at the word itself.
All sentences were in the past tense for consistency but the meaning of the sentences were as varied as possible to prevent repetitiveness. See (9) for an example of target sentence used in the experiment. The remaining target sentences can be found in the appendix.

(9) Ik plantte **er** een paar erg grote
I planted there a few very large
bomen **naast** voor een mooier
trees next for a nicer
totaalplaatje.
total picture.
'I planted a few very large trees next to it for a nicer overall picture'

#### 2.5.2 Baseline

Baseline sentences were the exact same sentences as the target sentences, but had $er_P$ removed and the object that $er_P$ referred to put in after the preposition. With these sentences, we can compare the time needed to read the preposition and see what kind of influence $er_P$ has on this. These sentences are very similar to the target sentences and therefore the target and baseline version of the same sentence are not both shown to a single participant. See (10) for the baseline version of (9), the rest can again be found in the appendix.

(10) Ik plantte een paar erg grote bomen
I planted a few very large trees
**naast** *het huis* voor een mooier
next the house for a
totaalplaatje.
nicer total picture.
'I planted a few very large trees next to the house for a nicer overall picture'

#### 2.5.3 Filler

Filler sentences did not have any specific grammatical structure or specific meaning but were designed to keep the participant from noticing the grammatical effect of the target sentences. The sentences were on average equal in length in comparison to the target and baseline sentences and did not contain 'er'.

### 2.6 Region of Interest

Within the target sentences, we are interested in a specific Region Of Interest (ROI) where we expect to see the effect we hypothesized. because we want to see the effect of $er_P$ on reading speed and we know from previous literature that LDDs show a slowdown when resolving the LDD, the ROI for this study is the preposition, the second part of the LDD. To account for the spillover effect, the two words after the preposition are also used for the ROI.
Because this is an eye tracking experiment and even though calibration was done before the experiment and a fixation dot was shown between sentences to correct for any drifting, the coordinates in the data might still be off by a few pixels in comparison to where the participant is actually looking. It also does not contain the size of the focus of what we can read when looking at a single point. To account for this each boundary of the ROI is padded by 10 pixels.

### 2.7 Data Collection

For each participant the following data was collected: First is which sentences were shown to what participant and in which order. Second is the data from the experiment that contained where on a screen a specific word was shown.Third is the eye tracking data that contained for each trial all the data the eye tracker had logged. Next to this a file containing all the sentences in correct order and a file containing in which order participants were shown the sentences were created to help with data analysis.

### 2.8 Data Processing

Eye tracking data is very big and convoluted. To make the data more usable, we subtracted for each fixation of a participant, which trial and fixation it was, its start, end and duration and the coordinates of the fixation. This was then linked to the correct sentence to show where participants looked and how long for each sentence.
From this, two data sets were created. The first one contained for each sentence and participant, the first fixation at the preposition and the first fixation after the preposition. the second data set, just as the first one, contained for each sentence and participant the first fixation at the preposition, but also contained the first fixation at two words after the preposition as well.All data analysis was done in Rstudio, version 4.1.2, found **here**.

# 3 Results

To see if $er_P$ had an effect on reading speed, a self paced reading experiment in Dutch was created. Each of the 26 participants read 40 sentences of which 10 were target, 10 were baseline and 20 were filler. This leaves us with 260 target sentences, 260 baseline sentences and 520 filler sentences. In total there were 20 different target and 20 different baseline sentences with equal frequency giving us 13 samples per sentence. The filler sentences were not used in analysis since they did not contain the effect this study was trying to measure.
As said in the methods section, the eye tracking data, which was an .edf file that contained a detailed record of everything that happened between the sentence showing and the pressing of the space bar. This data was extracted into a .csv file containing the following: for every fixation in a sentence, the number of the fixation was taken together with the start and end times of the fixation with relation to the start of the sentence, the duration of the fixation and the X and Y coordinates of the fixation.
This was combined with the experiment data that contained the boundaries for where the words were shown on the screen. Each fixation was linked to the correct sentence and participant to create a single data set containing, for each sentence and participant, the fixation number, start end and duration of the fixation and the X and Y coordinates of the fixation. All of this was done using a script.

To analyze if $er_P$ had an effect on the reading times of the preposition the two data sets as described in section 2.8 were created. The second dataset showed that there are a small number of trials with very high reading times in comparison to the rest of the data. It is likely that these trials were outliers where the participant got distracted by something in the environment and are therefore excluded from the analysis. The cut-off point was chosen to be 1.5 times the Inter Quartile Range (IQR), giving us 2.7 standard deviations of data to both sides of the mean. Any trials that fall outside this range were not used in analysis. Figure 3.1 and figure 3.2 show the distribution of reading times for a preposition and the two following words for target and baseline sentences respectively. All data above the red lines are not used in analysis.

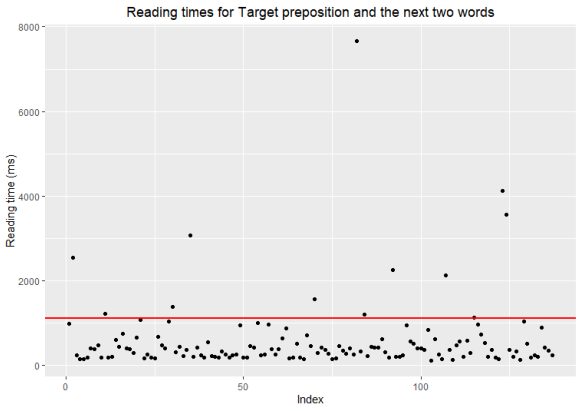


**Figure 3.1: The reading times for all target sentences and all participants with the spillover region. All data points above the red line are considered outliers based on 1.5*IQR**

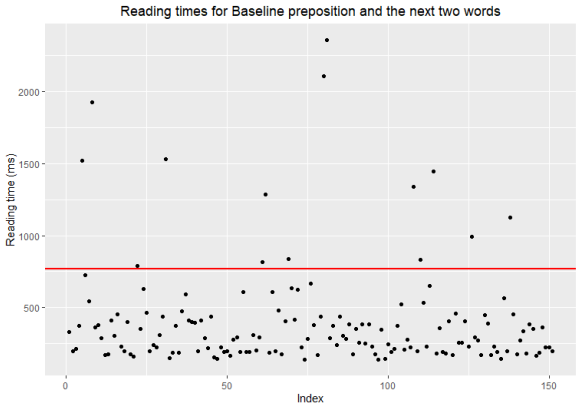


**Figure 3.2: The reading times for all baseline sentences and all participants with the spillover region. All data points above the red line are considered outliers based on 1.5*IQR**

The difference in reading times between target and baseline sentences for the preposition was analysed. This would indicate the time needed to process the preposition. The first dataset that contained, for each sentence and participant, the first fixation at the preposition and the first fixation after the preposition was used. The reading times are converted to a logarithmic scale since the data was not normally distributed otherwise, see figure 3.3 and 3.4. Both the groups are very similar in distri-

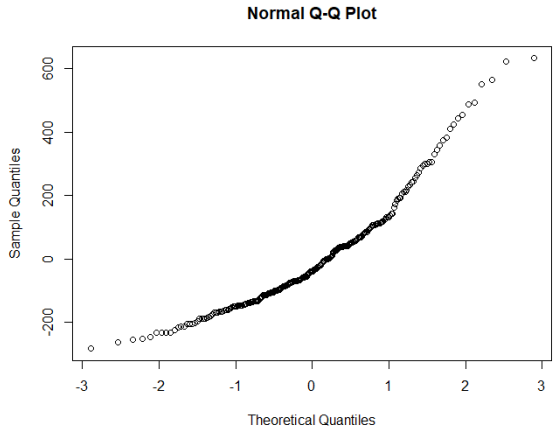


**Figure 3.3: A Q-Q plot showing the distribution of residuals over the quantiles for the normal reading times.**

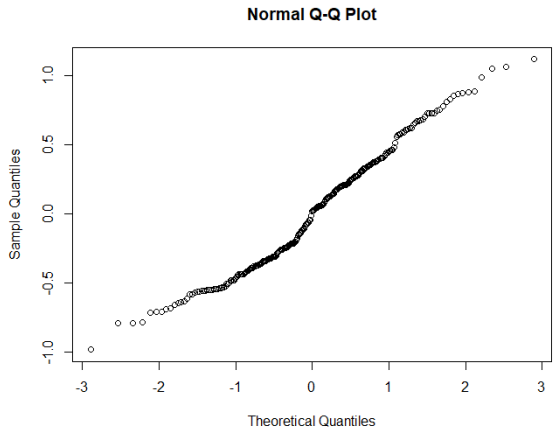


**Figure 3.4: A Q-Q plot showing the distribution of residuals over the quantiles for the logarithmic reading times.**


**Figure 3.5: A violin and box plot showing the distribution of reading times of the preposition for the target and baseline sentences on a logarithmic scale. The mean for each group is shown as a red dot and a line between the two is shown for better comparison.**

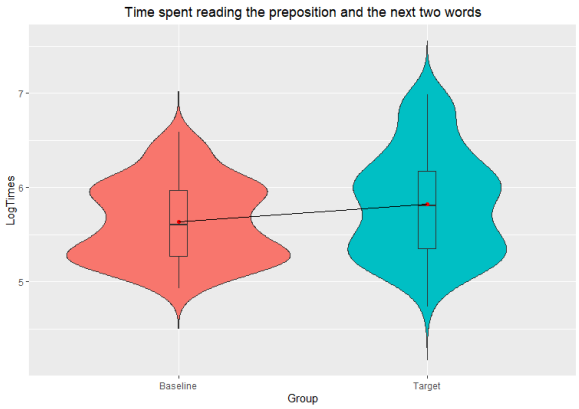


**Figure 3.6: A violin and box plot showing the distribution of reading times of the preposition plus two words for the target and baseline sentences on a logarithmic scale. The mean for each group is shown as a red dot and a line between the two is shown for better comparison.**

bution and mean, indicating that there is no difference, see figure 3.5.

Next, the second data set was used that included an expected spillover effect where the processing of the words happens a little bit later than when the eyes are looking which was the second dataset. As explained earlier, this was done by taking the first fixation at the preposition and the first fixation at two words after the preposition. The difference is the time needed to read this part of the sentence. This showed that on average, target sentences took longer to read than the corresponding baseline version of the sentence, see figure 3.6 To see if the difference in reading speed between the target and baseline sentences when accounting for a spillover effect is significant, a mixed model was created. This mixed model had the group as the

```
> summary(stat.model)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: LogTimes ~ Group + (1 | Participant) + (1 | Sentence)
   Data: statData

     AIC      BIC   logLik deviance df.resid
   366.8    384.6   -178.4    356.8      257

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.15905 -0.79707  0.04147  0.71097  2.44567

Random effects:
 Groups      Name        Variance Std.Dev.
 Sentence    (Intercept) 0.006842 0.08272
 Participant (Intercept) 0.020893 0.14455
 Residual                0.207470 0.45549
Number of obs: 262, groups:  Sentence, 40; Participant, 26

Fixed effects:
            Estimate Std. Error t value
(Intercept)  5.64283    0.05228 107.931
GroupTarget  0.17148    0.06300   2.722

Correlation of Fixed Effects:
            (Intr)
GroupTarget -0.579
```

**Figure 3.7: The R output for the summary of the mixed model.**

fixed effect with the different participants and sentences as random intercepts. An ANOVA on the model in comparison to a NULL model that did not account for the difference between target and baseline showed that the target group affected the log reading times ($\chi^2(1)$=6.84, p=0.0089), increasing it by about $0.17 \pm 0.06$ (SE). The R output that created these results can be found in figure 3.7 and 3.8. In 3.7 we can see the summary of the mixed model with the formula used and the quality of the statistical model using different criteria. The distribution of the residuals is shown using quantiles. The random effects are shown with most interestingly a residual effect of 0.207, a lot bigger than the random effects explained by the variables used in the model. the fixed effects show the influence of the target sentences on the reading times on a logarithmic scale.
In 3.8 we see the comparison between the mixed model and a NULL model that did not account for a difference between target and baseline. Here we see again the formula used for the models together with the quality of both models. The mixed model also shows the p-value for comparing the the two models.

```
> anova(stat.null,stat.model)
Data: statData
Models:
stat.null: LogTimes ~ (1 | Participant) + (1 | Sentence)
stat.model: LogTimes ~ Group + (1 | Participant) + (1 | Sentence)
           npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
stat.null     4 371.63 385.90 -181.81   363.63
stat.model    5 366.79 384.63 -178.40   356.79 6.8364  1   0.008932 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

**Figure 3.8: The ANOVA comparing the mixed model to a NULL version that did not account for the difference between target and baseline.**

# 4 Conclusion & Discussion

## 4.1 Conclusion

This paper investigated the effect of the grammatical construction $er_P$ on reading speed in a Dutch reading experiment. By doing a statistical analysis on a mixed model using an ANOVA, a significant increase has been found in the reading times of the preposition linked to $er_P$ in comparison to the preposition without an $er_P$ attached to it. From this, we can conclude that there is indeed an effect of $er_P$ on reading speed and that it takes longer to read a sentence with $er_P$ than the same sentence without $er_P$. This is in line with what was hypothesized that $er_P$ would increase reading times at the preposition since the interpretation of $er_P$ is left open until the preposition is read and processed. This means that the effect is only visible when the preposition has been processed, which happens in the spillover region instead of at the preposition.
The findings of this study suggest that the grammatical structure $er_P$ indeed seems to function like a Long Distance Dependency and are confirmatory to the results found in Stowe (1986) where there was a noticeable slowdown when resolving the LDD. From this study it can also be concluded that LDDs exist in the Dutch language and that they behave in the same way that WH-questions do in English as used by Stowe (1986).
Based on previous research by Grondelaers et al. (2009) and Grondelaers (2020), we know that other functions of ‘er’ decrease reading times by acting as an expectancy monitor for unexpected subjects. This study shows that not all functions of ‘er’ have the same effect and therefore behave differently depending on its function and has different effects on reading speed and comprehension.

## 4.2 Discussion

While the finding that $er_P$ takes longer to read than its baseline is significant, there are some potential flaws that should be mentioned.
First of all is that eye tracking data, while giving you a good insight into where someone is looking for how long, is not always entirely accurate and error prone. The width of a gaze is hard to track and since experiment uses word boundaries to order the fixations, the results may differ when using a different eye tracker or sampling rate. However, since there is not a way to follow an eye in real time, these flaws are not avoidable and are present in a lot of studies using eye tracking.
The second potential flaw is the way the sentences were structured. In the baseline sentences, $er_P$ was replaced with the subject it referred to and put after the preposition, see appendix. This is however also the spillover region used in the analysis for measuring the reading time of the preposition. This means that the spillover region for the target and baseline sentences is different, which might have an effect on the reading speed in that region. The level of effect probably changes with the complexity of the subject with more complex and longer subjects decreasing the gap in reading times and simple subjects increasing the gap in reading times. Changing the level of expectancy of the subject i.e. more or less fitting subjects might also change the reading speed gap between the target and baseline sentences, but this would have to be researched more. A third potential flaw is that some unconsidered confounding variable could impact the results. While all sentences were constructed carefully as explained in 2.5.1, some unconsidered grammatical construction might impact the results.
Now, in light of all these results and flaws, could they also mean something different than the effect this study tries to show? Since the effect is significant it can be assumed that $er_P$ indeed does increase the reading time of the preposition. However, as seen in figure 3.7, we can see that only 12% of the total variance is explained by the variables in the model. This could indicate that there are other variables that are influencing this difference that were not included in the analysis, which could bring to light some more interesting findings next to the one already shown in this study.

## 4.3 Future Research

To better know the significance of the findings of this study, more research needs to be done. For example, an experiment studying the general effect of LDD's on reading speed in Dutch would give us a better baseline to compare to and see if the effect found by Grondelaers et al. (2009) and Grondelaers (2020) that some functions of 'er' increase reading speed also hold true within LDD's.
A second interesting experiment would be to look at the effect of the distance between $er_P$ and its preposition. This would show if the effect changes with the length between $er_P$ and its preposition and would help interpret if the effect is more based on 'er' or that it is a general LDD. This would show if the results from Gibson and Warren (2004), where it is theorized that LDDs become more costly to keep unresolved the longer they get,would also hold true for $er_P$ and its Long Distance Dependency function. It is also possible that the length of the LDD does not matter and that the effect of $er_P$ stays the same no matter the length.
Lastly, a study varying the subjects that replace $er_P$ in the baseline sentences could give us more insight into the role the subject plays in the reading speed of the preposition plus spillover region. Using the similar sentences with $er_P$ but only changing the subject that replaces it in the baseline version could then be compared to the findings of this study. This could show that the reading times for the baseline sentences are also dependant on what kind of subject is chosen or that it is irrelevant for the effect, making the effect of $er_P$ more clear.

# References


Gunnar Bech. *Über das niederländische Adverbialpronomen er.* Nordisk sprog-og kulturforlag, 1952.

Hans Bennis. Gaps and dummies. *Dordrecht: Foris*, 1986.

Sanne Berends, Aafke Hulk, and Petra Sleeman. The emergence of the pronouns dutch er and french en in child l1 and the role of complexity. *Language Sciences*, 60:144–159, 2017.

Gosse Bouma. Finding long-distance dependencies in the lassy corpus. *Crossroads Semantics: Com-*

*putation, experiment and grammar*, pages 39–56, 2017.

Bruce Donaldson. *Dutch: A comprehensive grammar*. Routledge, 2008.

Lyn Frazier. Syntactic processing: evidence from dutch. *Natural Language & Linguistic Theory*, 5 (4):519–559, 1987.

Edward Gibson and Tessa Warren. Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, 7(1):55–78, 2004.

Stefan Grondelaers. Verwacht het onverwachte: Over hoe nederlanders er anders gebruiken dan vlamingen. *Nederlandse Taalkunde*, 25(2-3):205–211, 2020.

Stefan Grondelaers, Dirk Speelman, Denis Drieghe, Marc Brysbaert, and Dirk Geeraerts. 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, 2009.

Stephen M Jones. Multifunctional dutch 'er'. 2020.

Ronald M Kaplan and Annie Zaenen. Long-distance dependencies, constituent structure, and functional uncertainty. *Alternative conceptions of phrase structure*, 17:42, 1989.

Wim G. Klooster. Over zo'n and zo meer [on zo'n and so on]. in freek van de velde, hans smessaert, frank van eynde & sara verbru, 255-270, 2014.

Ad Neeleman and Hans Van de Koot. Syntactic haplology. *The Wiley Blackwell Companion to Syntax, Second Edition*, pages 1–31, 2017.

Johannes Engelbertus Josephus Maria Odijk. *Compositionality and syntactic generalizations*. PhD thesis, Katholieke Universiteit Brabant, 1993.

Colin Phillips, Nina Kazanina, and Shani H Abada. Erp effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22(3):407–428, 2005.

Laurie A Stowe. Parsing wh-constructions: Evidence for on-line gap location. *Language and cognitive processes*, 1(3):227–245, 1986.

Henk Van Riemsdijk. *A case study in syntactic markedness*. De Gruyter, 1978.

Gert Webelhuth and Olivier Bonami. Syntactic haplology and the dutch proform "er". In *Proceedings of the 26th International Conference on Head-Driven Phrase Structure Grammar*, pages 100–119, 2019.

# A Appendix

## A.1 Target Sentences

(1) Ik plantte er een paar erg grote bomen naast voor een mooier totaalplaatje.
I planted it a few very large trees next for a nicer overall picture.
'I planted a few very large trees next to it for a nicer overall picture.'

(2) Hij voer er met een paar extra minuten in zonder iets aan te raken.
He entered it with a few extra minutes in without something at to hit.
'He entered it with a few extra minutes without hitting something.'

(3) Ze maakte er heel vrolijk een grote bende van met haar vriendin.
She made it very cheerful a big mess from with her girlfriend.
'Very cheerfully she made a big mess from it with her girlfriend.'

(4) Zij deden er al meerdere maanden zonder zorgen mee toen het kapot ging.
They did it already multiple months without worries with when it broke.
'They already spent multiple months without worries with it when it broke. '

(5) Hij reed er met maar een beetje speling voor maar miste de brievenbus net.
He drove it with just a little bit of playroom in front but missed the letterbox just.
' He drove with just a little bit of playroom in front of it but just missed the letterbox. '

(6) Ik waste er elke dag mijn handen in bij gebrek aan beter.
I washed it every day my hands in with lack of better.
'I washed my hands in it since there was no better option. '

(7) De man luisterde er elke dag 's morgens vroeg naar om de dag goed te beginnen.
The man listened it every day in the morning early to to the day well to start.
' The man listened every early morning to it to start the day well.'

(8) De vrouw dronk er grote gulzige slokken uit toen ze proefde wat er in zat.
The woman drank it big greedy gulps out when she tasted what there in sat.
' The woman drank big greedy gulps from it when she tasted what was in it '

(9) Hij racete er bijna elke dag 's ochtends vroeg voorbij op weg naar Amsterdam.
He raced it almost every day in the morning early past on the way to Amsterdam.
' He raced almost every day in the early morning past it on the way to Amsterdam. '

(10) Ik stuurde er een heel lief en vriendelijk kaartje naartoe om haar te troosten.
I sent it a very kind and friendly card to to her to comforting.
' I sent a very kind and friendly card to her to comfort her. '

(11) Hij keek er heel verbaasd en respectvol naar met opengevallen mond.
He looked it very surprised and full of respect to with open mouth.
' He looked very surprised and full of respect to it with his mouth fallen open.'

(12) De student studeerde er elke dag zes uren lang voor om het tentamen te halen.
The student studied it every day six hours long for to the exam to pass.
' The student studied every day for six hours for it to pass the exam. '

(13) Ze brak er een redelijk groot stukje van af en gaf dat aan haar vriendje.
She broke it a decently big piece from and gave that to her friend.
' She broke a decently sized piece from it and gave that to her friend. '

(14) Het kind speelde er dagenlang met grote gretigheid in voordat het saai werd.
The kid played it days long with big eagerness in before it boring became.
‘ The kid played for days very eagerly in it before it became boring. ’

(15) De agent floot er met zijn politie fluitje naar om de bus te doen stoppen.
The policeman whistled it with his police whistle to to the bus it make stop.
‘ The policeman whistled with his police whistle to the bus to make it stop. ’

(16) De sporter klom er met geoefende behendigheid zo in zonder snelheid te verliezen.
The sportsman climbed it with practised agility fast in without speed to lose.
‘ The sportsman climbed with practised agility fast in it without losing speed. ’

(17) Ik sliep er al wekenlang slecht en chagrijnig op voordat ik een nieuwe kocht.
I slept it already weeks long bad and grumpy on before I a new one bought.
‘ I slept bad and grumpy on it for weeks before I bought a new one. ’

(18) De vrouw gleed er met een snelle vaart in toen ze zich over de glijbaan liet gaan.
The woman slid it with a fast speed in when she herself over the slide let go.
‘ The woman slid very fast in it when she went on the slide. ’

(19) Hij schoot er met een paar millimeter naast en won de prijs niet.
He shot it with a few millimetres next and won the prize not.
‘ He shot next to it by a few millimetres and didn’t win the prize. ’

(20) Het kind kleurde er met grote precisie in zonder buiten de lijntjes te gaan.
The child coloured it with great precision in without outside the lines to go.
‘ The child coloured it in with great precision without going outside the lines. ’

## A.2 Baseline Sentences

(21) Ik plantte een paar erg grote bomen naast mijn huis voor een mooier totaalplaatje.
I planted a few very large trees next my house for a nicer overall picture.
’I planted a few very large trees next to my house for a nicer overall picture.’

(22) Hij voer met een paar extra minuten in de haven zonder iets aan te raken.
He entered with a few extra minutes in the port without something at to hit.
’He entered the port with a few extra minutes without hitting something’

(23) Ze maakte heel vrolijk een grote bende van haar kamer met haar vriendin.
She made very cheerful a big mess from her room with her girlfriend.
’Very cheerfully she made a big mess from her room with her girlfriend. ’

(24) Zij deden al meerdere maanden zonder zorgen met de auto toen hij kapot ging.
They did already multiple months without worries with the car when it broke.
’They already spent multiple months without worries with the car when it broke. ’

(25) Hij reed met maar een beetje speling voor het huis maar miste de brievenbus
He drove with just a little bit of playroom in front the house but missed the letterbox
net.
just.
’ He drove with just a little bit of playroom in front of the house but just missed the letterbox. ’

(26) Ik waste elke dag mijn handen in het fonteintje bij gebrek aan beter.
I washed every day my hands in the fountain with lack of better.
’I washed my hands in the fountain since there was no better option. ’

(27) De man luisterde elke dag ‘s morgens vroeg naar de podcast om de dag goed te
The man listened every day in the morning early to the podcast to the day well to
beginnen.
start.

’ The man listened every early morning to the podcast to start the day well.’

(28) De vrouw dronk grote gulzige slokken uit de beker toen ze proefde wat er in zat.
The woman drank big greedy gulps out the cup when she tasted what there in sat.

’The woman drank big greedy gulps from the cup when she tasted what was in it ’

(29) Hij racete bijna elke dag ‘s ochtends vroeg voorbij het monument op weg naar
He raced almost every day in the morning early past the monument on the way to
Amsterdam.
Amsterdam.

’He raced almost every day in the early morning past the monument on the way to Amsterdam.’

(30) Ik stuurde een heel lief en vriendelijk kaartje naar mijn oma om haar te troosten.
I sent a very kind and friendly card to my grandma to her comforting.

’I sent a very kind and friendly card to my grandma to comfort her.’

(31) Hij keek heel verbaasd en respectvol naar de paus met opengevallen mond.
He looked very surprised and full of respect to the Pope with open mouth.

’He looked very surprised and full of respect to the Pope with his mouth fallen open.’

(32) De student studeerde elke dag zes uren lang voor de stof om het tentamen te halen.
The student studied every day six hours long for the material to the exam to pass.

’The student studied every day for six hours for the material to pass the exam. ’

(33) Ze brak een redelijk groot stukje van de reep chocolade af en gaf dat aan haar
She broke a decently big piece from the bar of chocolate off and gave that to her
vriendje.
friend.

’She broke a decently sized piece from the bar of chocolate and gave that to her friend. ’

(34) Het kind speelde dagenlang met grote gretigheid in het huisje voordat het saai werd.
The kid played days long with big eagerness in the house before it boring became.

’The kid played for days very eagerly in the house before it became boring. ’

(35) De agent floot met zijn politie fluitje naar de bus om hem te doen stoppen.
The policeman whistled with his police whistle to the bus to him to make stop.

’The policeman whistled with his police whistle to the bus to make it stop.’

(36) De sporter klom met geoefende behendigheid zo in de boom zonder snelheid te
The sportsman climbed with practised agility fast in the tree without speed to
verliezen.
lose.

’The sportsman climbed with practised agility fast in the tree without losing speed. ’

(37) Ik sliep al wekenlang slecht en chagrijnig op mijn matras voordat ik een nieuwe
I slept already weeks long bad and grumpy on my mattress before I a new one
kocht.
bought.

’ I slept bad and grumpy on my mattress for weeks before I bought a new one. ’

(38) De vrouw gleed met een snelle vaart in het zwembad toen ze zich over de
The woman slid with a fast speed in the swimming pool when she herself over the
glijbaan liet gaan.
slide let go.
'The woman slid very fast in the swimming pool when she went on the slide. '

(39) Hij schoot met een paar millimeter naast de roos en won de prijs niet.
He shot with a few millimetres next the bullseye and won the prize not.
' He shot next to the bullseye by a few millimetres and didn't win the prize.'

(40) Het kind kleurde met grote precisie in het tekenboek zonder buiten de lijntjes te
The child coloured with great precision in the drawing book without outside the lines to
gaan.
go.
'The child coloured the drawing book in with great precision without going outside the lines.'

## A.3 Filler Sentences

(41) Vandaag kwam het pakketje aan waar ik al bijna vier weken op gewacht had.

'Today the package I had been waiting for almost four weeks arrived.'

(42) Ik wilde wel naar de wedstrijd toegaan, maar door mijn klachten moest ik thuisblijven.

'I wanted to go to the game, but because of my symptoms I had to stay at home.'

(43) Als het straks gaat regenen kunnen we misschien beter nu al met de bus gaan.

'If it starts to rain later, we might be better off taking the bus now.'

(44) Ze vertelde me vandaag dat haar laatste examen toch wel moeilijker was dan verwacht.

'She told me today that her last exam was more difficult than expected.'

(45) Hij keek om naar zijn ouders en zag dat ze beide tranen in hun ogen hadden.

'He looked back at his parents and saw that they both had tears in their eyes.'

(46) Vanaf morgen wordt de kans steeds groter dat de temperatuur onder de nul zal blijven.
hoi
'From tomorrow there is a growing chance that the temperature will remain below zero.'

(47) Altijd wanneer ik ergens op tijd wil zijn komt er weer iets tussen waardoor ik te laat ben.

'Whenever I want to be somewhere on time, something comes up that makes me late.'

(48) De man voelde zich erg beledigd toen ik zei dat hij er oud uitzag.

'The man was very offended when I said he looked old.'

(49) Een krantenwijk is vaak het eerste baantje van tieners, maar niet iedereen houdt het lang vol.

'A newspaper route is often the first job of teenagers, but not everyone lasts long.'

(50) Volgende week komt er een monteur langs die onze nieuwe CV-ketel zal aansluiten.

'Next week a mechanic will come by who will connect our new central heating boiler.'

(51) Zonder te kijken kon ik alles opnoemen wat op de menukaart van het restaurant stond.

'Without looking, I could name everything on the restaurant menu.'

(52) Het voetbalstadion barstte in gejuich uit toen de thuispartij een doelpunt maakte.

'The football stadium erupted in cheers when the home side scored a goal.'

(53) Morgen is er weer een dag, zei ik nadat ik vandaag weer niks gedaan had.

'Tomorrow is another day, I said after I had done nothing again today.'

(54) Een bezoekje bij oma voelt altijd goed, je kan echt zien dat ze ervan geniet.

'A visit to grandma always feels good, you can really see that she enjoys it.'

(55) Zonet ging alles nog goed maar nu lijkt het alsof het nooit meer goed zal gaan.

'Everything was going well just now, but now it seems like it will never go well again.'

(56) Soms is het beter om even niks te zeggen en er gewoon voor iemand te zijn.

'Sometimes it's better to say nothing and just be there for someone.'

(57) Het jongetje huilde toen hij van de fiets viel en zijn knie schaafde.

'The boy cried when he fell off the bicycle and scraped his knee.'

(58) Luisterend naar de professor leerde de student steeds meer over het onderwerp.

'Listening to the professor, the student learned more and more about the subject.'

(59) In het vliegtuig kon ik niet slapen door het kletsende echtpaar dat naast me zat.

'I couldn't sleep on the plane because of the chatting couple sitting next to me.'

(60) Bij het tellen van de stembiljetten waren dit keer wel erg veel fouten gemaakt.

'In the counting of the ballots, a lot of mistakes had been made this time.'

(61) Nadat ik de trein uit gestapt was bedacht ik me dat mijn tas nog in de trein lag.

'After I got off the train I realized that my bag was still in the train.'

(62) Door de werkzaamheden kon ik vandaag alweer niet op tijd op werk zijn.

'Due to the construction work I was unable to get to work on time again today.'

(63) Het is al lang geleden dat ik voor het laatst in de bioscoop geweest ben.

'It's been a long time since I last went to the cinema.'

(64) Ze eet elke dag havermout als ontbijt voordat ze naar school moet.

'She eats oatmeal for breakfast every day before going to school.'

(65) Voor een bruiloft moet je jezelf netjes kleden, alledaagse kleding is dan niet oké.

'For a wedding you have to dress yourself neatly, everyday clothes are not okay.'

(66) In het park bloeien de bloemen en hoor je de vogeltjes fluiten.

'The flowers bloom in the park and you can hear the birds chirping.'

(67) Hij is altijd tevreden over zijn kapper, hij wordt altijd precies goed geknipt.

'He is always satisfied with his hairdresser, he is always cut just right.'

(68) Vanaf morgen kun je weer stoofpeertjes en kersen in de supermarkt halen.

'Starting tomorrow you can get stewed pears and cherries in the supermarket again.'

(69) Om vier uur 's middags komt even een schoonmaker langs om de troep van gister op te ruimen.

'At four o'clock in the afternoon a cleaner comes by to clean up yesterday's mess.'

(70) Etensresten kun je in de koelkast maar een paar dagen bewaren, maar in de vriezer veel langer.

'Food leftovers can only be kept in the fridge for a few days, but much longer in the freezer.'

(71) De busreis naar Madrid duurde acht uur langer dan verwacht door een ongeluk.

'The bus trip to Madrid took eight hours longer than expected due to an accident.'

(72) Dierentuinen hebben heel veel verschillende soorten dieren, maar de stokstaartjes vind ik het leukst.

'Zoos have many different kinds of animals, but the meerkats are my favorite.'

(73) Over drie weken kun je eindelijk tickets bestellen voor het concert in maart.

'In three weeks you can finally order tickets for the concert in March.'

(74) Bij het restaurant waren de vrienden heerlijk aan het dineren voor een laag prijsje.

'At the restaurant the friends were having a delicious dinner for a low price.'

(75) 's Nachts word ik elke keer wakker van mensen die over straat schreeuwen.

'At night I wake up every time to people shouting in the street.'

(76) Op school leer je allemaal dingen die later in je leven nodig gaat hebben.

'At school you learn all kinds of things that you will need later in life.'

(77) Bij het café krijg je de tweede koffie gratis als je voor twaalf uur langskomt.

'At the cafe you get the second coffee for free if you come before twelve o'clock.'

(78) Vrijdag vind ik de leukste dag van de week want dan begint het weekend.

'Friday is my favorite day of the week because that's when the weekend starts.'

(79) Als het morgen alweer niet doorgaat, dan zeg ik het helemaal af.

'If it doesn't happen tomorrow, I'll cancel it completely.'

(80) Zonder te kijken sloeg hij linksaf en fietste bijna een voetganger aan.

'Without looking, he turned left and nearly hit a pedestrian.'