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

The Effect of Prepositional *er* on Processing Speed: an Empirical Examination

Bachelor's Project Thesis

Joas Mol, s4641353, j.mol.8@student.rug.nl,
Supervisor: Dr S.M. Jones

Abstract: This project aims to expand the understanding of the effect of long-distance dependencies (LDD's) on word processing speed in Dutch. It has been found that LDD's in *what*-questions increase processing load at the resolving point (Badecker & Straub, 2002). However, with respect to the LDD's caused by the different forms of the Dutch word *er*, an empirical observation of this effect has not been made. In this project, we examined the effect of the long-distance dependency of prepositional er on processing load. We used a moving window self-paced reading task to compare the reading speeds of sentences with er_p, sentences with wh-constructions, and sentences without LDD's. We have found no significant difference between any of the conditions. This means we have observed no slowing effect of resolving an LDD in sentences with er_P, nor have we observed a previously found slowing effect of resolving an LDD in *what*-questions. However, due to likely methodological issues, we believe the data to be an unreliable indicator of this effect.

1 Introduction

Cognitive Linguistics as a research field is concerned with the intersection between language and cognition (Robinson & Ellis, 2008). It stems from – and is closely related to – Classical Linguistics and Cognitive Psychology. It studies the processes and mechanisms of language in the human brain and creates linguistic models based on these cognitive data.

One of the most common cognitive features Cognitive Linguistics tries to study is processing load. Processing load is a measure of how much effort the brain needs to process a linguistic unit (e.g. words, word groups, sentences). There are multiple factors of words that are known to influence processing load. The most influential factor is word frequency (also called commonality) (Brysbaert et al., 2018; Baayen et al., 2016): words used frequently are processed more easily than words used very infrequently. *Expectation* also has an effect on processing load (Grondelaers et al., 2002): words that are expected within the semantic context are processed more easily than words that are unexpected. As such, 1a is processed more easily than both 1b and 1c.

(1) a. On my roof, there's a bird.

b. On my roof, there's a budgerigar.

c. On my roof, there's a penguin.

However, since processing load cannot be directly measured, we must observe a consequence of it instead. The most powerful option is to directly measure brain activity. However, this is also the most complex to implement and interpret, and is often not necessary. Therefore, we will avoid using this if possible. This leaves us with two comparatively similar options: examining eye movement and reading speed.

Eye movement can be measured with an eye-tracking paradigm ("Eye-Tracking with Text", 2014). During an eye-tracking experiment, participants are tasked to read a stimulus shown on a computer screen. During the trial, the eyes fixate on different words. This *fixation point* is recorded over time. Fixation duration has been linked to processing speed (Just & Carpenter, 1980), so the processing speed of individual words can be determined from the fixation data.

Reading speed can also be examined using a self-paced reading (SPR) paradigm (Jegerski, 2014).

During an SPR experiment – specifically the *moving window* variant – participants are shown a stimulus in stages, as in Figure 1.1. At any point, one singular word is shown and all other words are masked. Participants are instructed to press a key whenever they are done processing the current word. This key press progresses the trial to the next word in the stimulus. Participants are instructed to process words before continuing, so the processing speed of individual words can be determined from the response time data.

Figure 1.1: An example of a moving window self-paced reading trial. Each x represents a key-press

For reading experiments, one must take into account *processing load spillover* (Vasishth, 2006; Findelsberger et al., 2019). When words with a high processing load appear in a sentence, subsequent words take longer to process. This indicates that word processing is not fully finished when a person stops attending to it. The effect diminishes as temporal distance to the high-load word increases. As such, the confounding effects of processing load spillover can be negated by increasing the distance between the region of interest and high-load words.

1.1 Long-distance dependencies

In this paper, we will discuss and examine *long-distance dependencies* (LDD). An LDD occurs when one word or word group in a sentence fills a *syntactic gap* in another arbitrary part of the sentence. In Ex. 2 we see a demonstration. Sentence 2a contains a subject, a verb and an object. We see in sentence 2b that when this sentences is converted into a *what*-question, the object is omitted. That is, *what* fills a syntactic role necessary for a grammatical sentence, which would normally be filled by the object. This location is marked with a — for clarity. We see in 2c that if both *what* and the object are omitted, the resulting question sentence is only dubiously grammatical and has a different meaning and contains no LDD.

(2) a. He receives a cat.

b. What$_i$ does he receive —$_i$?

c. ? Does he receive?

There are also other cases in which LDD's appear in English, but they are not relevant for this study. More importantly, we see LDD's for *wh-questions* in different languages. Both German and Dutch form LDD's when converting a simple sentence into a *wh*-question, illustrated in 3 and 4 respectively. In this simple example, the resulting word order is very similar to English, but we do see that both Dutch and German swap the subject and verb, and English adds a form of *to be*.

(3) a. *Hij krijgt een kat*
He receives a cat

'He receives a cat.'

b. *Wat krijgt hij —?*
What receives he ?

'What does he receive?'

(4) a. *Er bekommt eine Katze*
He receives a cat

'He receives a cat.'

b. *Was bekommt er —?*
What receives he ?

'What does he receive?'

When reading a sentence with a *wh*-LDD, the LDD is resolved at the point just before the syntactic gap. An increase in processing load has been found at this resolving point Badecker & Straub (2002).

1.2 Er

In addition to the LDD in *wh*-questions, the Dutch language also contains LDD's with the word *er*. *Er* is a pronoun that has a variety of unique syntactic rules and interactions. It is often equivalent to the English '*there*'. Grondelaers et al. (2009) found that *er* reduces the processing load increase for *unexpected* words. *Er* has four distinct roles in a sentence, which are classified into different forms. There are many classifications and specific namings for these forms, for this paper we will be using the classification of Odijk (1993).

1. Existential *er*
2. Locative *er*
3. Quantitative *er*
4. Prepositional *er*

Existential er is commonly used when the subject is indefinite. We see in Ex. 5a the basic structure of a Dutch sentence, with a *definite* subject. When the subject is *indefinite*, the same structure is grammatical but not preferred (5b). In Ex. 5c we see what happens in common Dutch usage: the subject is moved forward in the sentence and *er* is placed in its original place. This form of *er* is the only one that can appear alone at the start of a sentence.

(5) a. *Het boek ligt op tafel.*
The book lays on table.
'The book is on the table.'

b. ? *Een boek ligt op tafel.*
A book lays on table.
'There is a book on the table.'

c. er_X *ligt een boek op tafel.*
There lays a book on table.
'There is a book on the table.'

Locative er can replace *daar* (there) when it is in an unstressed position. As it cannot be in a stressed position, it is not possible to replace the *daar* in (6a) with *er* (6b). We see in Ex. 7 that these variants have the same semantic meaning. Er_L refers to a location, generally physical.

(6) a. *daar koop ik een boek*
there buy I a book
'I buy a book there.'

b. * *er koop ik een boek*
there buy I a book
(Intended) 'I buy a book there.'

(7) a. *Ik koop daar een boek.*
I buy there a book.
'I buy a book there.'

b. *Ik koop* er_L *een boek.*
I buy there a book.
'I buy a book there.'

Quantitative er is paired with a number or quantity in a sentence. As Ex. 8 shows, er_Q replaces the object of the *verb*. With this, it produces an LDD. The er_Q fills the syntactic gap of *boeken* (8b), and removing it creates a sentence which is not grammatical with the same meaning (8c). The 'seventeen' in the sentence is no longer a quantifier of an object, but the object itself (and would only be grammatical if one was buying a physical object in the shape of '17').

(8) a. *ik heb zeventien boeken gekocht*
I have seventeen books bought
'I have bought seventeen books.'

b. *ik heb* er_Q *zeventien — gekocht*
I have there seventeen bought
'I have bought seventeen (of them).'

c. * *ik heb zeventien gekocht*
I have seventeen bought
(Intended) 'I have bought seventeen'

Prepositional er replaces the object of a *preposition*. Replacing the object of a preposition with a pronoun referring to a thing (not a person) is ungrammatical (9a,9b). Instead of the pronoun, an instance of er_P is added to the sentence, and the object of the preposition is left out entirely, creating a dependency between the two (9c). It is generally added directly following the verb, though there is some variability in this. If er_P occurs directly before a preposition, the two words are concatenated into one (9c). However, it is possible to insert an adjunct between er_P and its preposition, splitting them and showing that er_P can cause an LDD. In these cases dutch speakers prefer splitting the words over keeping them concatenated (9d, 9e).

Sentences with er_P can be converted into a *wh*-question. The question word will always be '*waar*' in this case (9f).

(9) a. *ik kan met de zaag omgaaan*
I can with the saw go.around
'I can handle the saw.'

b. * *ik kan met het omgaan*
I can with it go.around
(Intended) 'I can handle it.'

c. *ik kan ermee — omgaan*
I can with.it go.around
'I can handle it.'

d. ? *ik kan ermee — goed omgaan*
I can with.it good go.around
'I can handle it well.'

e. *ik kan er goed mee — omgaan*
I can it good with go.around
'I can handle it well.'

f. *waar kan ik goed mee — omgaan*
where can I good with go.around
'What can I handle well?'

While sentences with er_Q and er_P can exhibit the syntactic characteristics of LDD's, their empirical effects on reading speed have not been studied. As such, we do not know how they behave. Do they exhibit the same effect as *wh*-questions?

In this paper, we will examine how er_P affects processing speed in a sentence compared to other long distance dependencies in Dutch. For this, we will perform a self-paced reading task using sentences with er_P, *wh*-questions, and a control condition with neither. We will compare reading speed at the *preposition* whose object may be removed. We predict that er_P slows down reading speed when resolving its LDD. The control condition will be used to compare against the *er* condition, and the comparison between the *wh* and control condition will be used to test the consistency of our results with respect to previous research.

2 Methods

For this experiment, we will be using the moving window SPR paradigm to test our hypothesis. We chose this paradigm because while the large variety of data collected by eye-tracking is useful to experiments that need it, the confounding possibilities induced can increase the reliability of the data. Since our experiment does not need the high flexibility of an eye-tracking paradigm we therefore chose for an SPR experiment. This means that our collected data is limited to one Reaction Time (RT) per word and does not take into account the delayed processing of words seen before. Our region of interest (ROI) will be the place where the LDD is resolved. This is the place where a higher processing load is expected.

2.1 Participants

Nine participants completed the experiment. They were native Dutch speakers, were between 18 and 60 years old, and had normal to corrected vision. The average age of participants was 25.6 years. Participants signed an informed consent form prior to the experiment.

2.2 Materials

Participants sat in a room with no audiovisual distractions, 1.2 meters from a screen on which the stimuli were displayed.

Black text was projected on a white background with a text size of 28 px in the Consolas font.

The experimental software was written with the JavaScript package JSPsych (De Leeuw et al., 2023), with SPR module code adapted from code by Josh De Leeuw[*].

Participants were first shown instructions to the experiment, after which they completed one example trial. Then, they started the experiment.

[*] https://github.com/jspsych/tutorials/blob/master/moving-window/

2.3 Stimuli

To negate any possible effect of word frequency on processing speed, all stimulus words were taken from *A frequency dictionary of Dutch* (Donaldson, 2008). The book orders its words based on their frequency, giving a frequency coefficient to each one. Due to the difference in frequency of different Parts of Speech (PoS), we determined an independent acceptable frequency coefficient range for each PoS. The acceptable frequency coefficient ranges for every PoS is included in Appendix A.

The experiment was designed to test three different conditions: an *er*-condition, a wh-condition, and a control condition. These three conditions were structured so that their differences were minimal. Each stimulus was a Dutch sentence constructed according to a strict template structure.

ER	SUBJ	AUX	*er*	ADJ	(OBJ_V)	**PREP**		VERB
WH	*waar*	AUX	SUBJ	ADJ	(OBJ_V)	**PREP**		VERB
CTR	SUBJ	AUX		ADJ	(OBJ_V)	**PREP**	OBJ_P	VERB

Figure 2.1: The template for sentence structures of stimuli. ER is the *er*-condition, WH is the wh-condition, and CTR is the control condition

An overview of this structure can be seen in Figure 2.1. We will now discuss this template and the rules that formed it. The capitalised terms represent words with that PoS. The word '*er*' is an instance of er_P. As discussed, '*waar*' is the *wh* equivalent of er_P and will be its counterpart in the *wh* condition. Both of these terms are the antecedent to an LDD with the term PREP. The control condition does not contain an LDD, so it does not contain an always-occurring word.

The term that is central to this experiment is PREP. This is the preposition at which the LDD is resolved. As such, it is our ROI for this experiment. The control condition has a preposition without a syntactic gap. This gap is instead filled by OBJ_P, the object of the preposition. This preposition is the syntactic equivalent of the PREP in the other conditions and will as such be used for comparison. According to our hypothesis and previous research (Stowe, 1986; Crain & Fodor, 1985) respectively, we should see a reading slowdown at the PREP for the *er*- and wh-conditions, but not at the preposition in the control condition.

The term ADJ represents an adjunct phrase that is inserted into the sentence to mitigate the *processing load spillover* effect (Vasishth, 2006; Findelsberger et al., 2019). The adjunct ensures that the distance between possible high-load words is high enough that the spillover effect is negligible on experimentally relevant words. As such, the adjunct should not contain any prepositions, as this may mislead participants into thinking the LDD is being resolved in the adjunct.

The term *OBJ* is in parentheses. This is because there are two stimulus variants: an *Object* variant (including the OBJ_V) and a *No-Object* variant (excluding the OBJ_V). We do not expect to see a difference in effect between the Object and No-Object variants, but any difference will be identified if present.

Using these rules, we generated thirty triplets of similar sentences to use as stimuli. An example of a stimulus triplet, with colours matching the template to highlight relevant structures, can be found in Ex. 10.

(10) a. *De politie gaat er grondig onderzoek naar doen.*
The police goes there thorough investigation to do.
'The police will thoroughly investigate it.'

b. *Waar gaat de politie grondig onderzoek naar doen?*
Where goes the police thorough investigation to do?
'What will the police thoroughly investigate?'

c. *De politie gaat grondig onderzoek naar de gevolgen doen.*
The police goes thorough investigation to the consequences do.
'The police will thoroughly investigate the consequences.'

To present these stimuli, we employed a Latin Square design: from every triplet, one stimulus is randomly chosen to be presented to the participant. This creates a list of conditions called a *seed*. To ensure that all stimuli are presented equally, seeds are coordinated between three participants (Table 2.1). Each of the three participants gets presented a different condition, ensuring that every condition is used exactly once for every three participants. Each participant was presented thirty experimental stimuli.

	Experimental Condition					
Part. 1	er	wh	er	ctrl	wh	ctrl
Part. 2	wh	er	ctrl	er	ctrl	wh
Part. 3	ctrl	ctrl	wh	wh	er	er

Table 2.1: Example seed of distribution for six stimuli across three participants

Complete lists of all stimuli can be found in Appendices B–D.

2.4 Filler trials

While these rigid templates and distribution methods are experimentally useful, they are also relatively transparent. To prevent participants from deriving the experimental goal from the experiment structure, we took two connected measures.

Firstly, fifteen *filler stimuli* were included in the experiment. A majority of these fillers were ungrammatical, but otherwise had no intended pattern. Participants were not told whether a stimulus was a filler. For every two real stimuli, a filler was inserted *before*, *between*, or *after* the two stimuli. The exact placement was randomly determined for every filler independently.

Secondly, to make use of the ungrammaticality of the fillers and divert attention more, participants were asked after every stimulus whether it was grammatical or not. This was intended to direct the participants' attention to the grammaticality of the stimuli, rather than their structure.

Each participant was shown all filler stimuli, bringing the total amount of stimuli each participant was presented to fourty-five.

A full list of filler stimuli can be found in Appendix E.

3 Results

We collected RT data from every word in the stimuli, but we will specifically examine the ROI as described in Section 2.

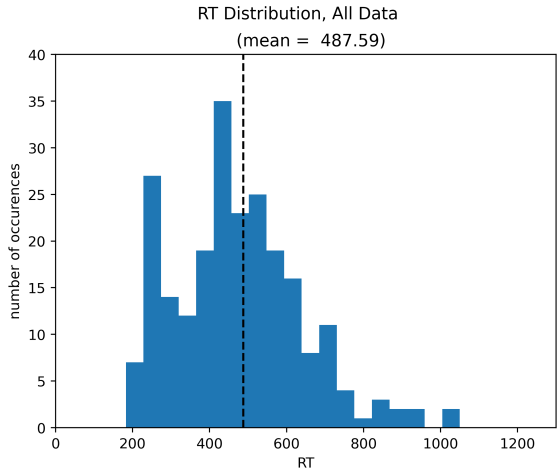

Figure 3.1: RT Distribution at LDD Resolution

The RT distribution can be seen in Fig. 3.1. The dataset contains a light positive skew, visible in the difference between the left and right tail. The left tail has a very steep slope, and the right tail is longer and shallower. The skewness coefficient (Pearson 2) of the dataset is 0.25. This indicates that the data is lightly skewed, but not enough to necessitate corrective transformation. We see that the distribution seems to be bimodal: it contains two peaks at 200ms and 400ms. This is expected if the conditions have a different mean RT.

3.1 Preprocessing

We divided these data into three *condition distributions*: one for each experimental condition (*er*, WH & control). We excluded data from four different stimuli (3, 6, 12 & 20) due to mistakes in stimulus design†. Additionally, individual outlier data points were removed. Points were considered outliers if they were more than 2 standard deviations away from the mean of the *condition distribution*. This resulted in seven more data points being removed from the dataset: two in the *er* condition, one in the WH condition, and four in the control condition. The outliers were distributed between multiple participants.

†These mistakes included not updating one condition when adding/removing a word, and in one case including a different stimulus for one of the conditions

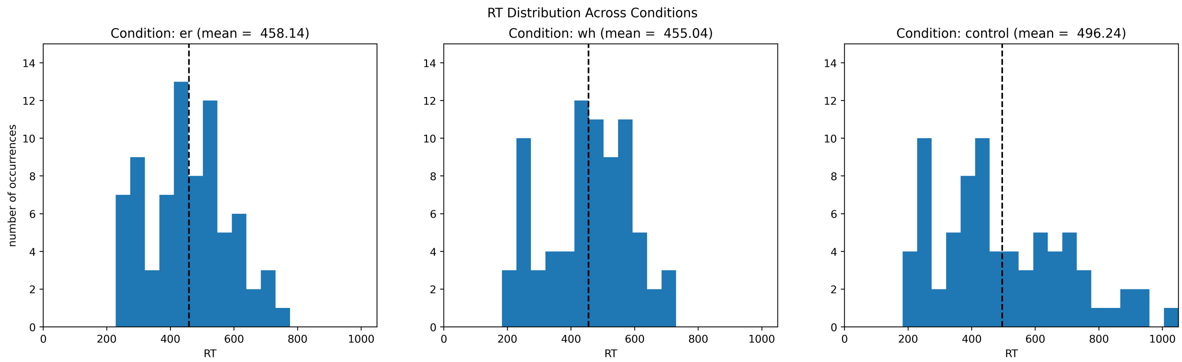

Figure 3.2: Reaction Time Distribution Across Conditions, Outliers Outside ±2SD Removed

3.2 Data distributions

However, when we examine the distribution across conditions, we see something unexpected (Fig. 3.2): the bimodality apparent in the full dataset is also present in the distributions of each condition individually. All three conditions contain a small peak around 200ms and another, larger peak around 400-500ms. We see in *control* distribution a long right tail that is absent in the other two distributions. We also see this in its skewness coefficient, which is a 0.65 compared to the *er* and *WH*-coefficients of 0.08 and -0.22 respectively.

We do see a difference in peak placement between the three conditions. The large peaks of both the *er* and WH conditions are both placed around 500ms, where the large peak of the control condition is placed at 400ms. This is in contrast to the means (marked with a black dotted line in Fig. 3.2), which is instead higher for the control condition. This discrepancy is likely due to the skewness of the control distribution.

We gain more insight into the key factors of these distributions by consulting Fig. 3.3, where we see clearly that while the range of the *er* and WH conditions is consistent, the control condition has a much larger right tail. This is not explained well by our hypothesis. On the contrary, our hypothesis would expect the right control tail to be shorter than those of the other two conditions.

We also see here compared the means and medians of the three conditions as the blue lines and white dots respectively. It is clear that while there is some variation between the three, there is no large visible effect between the *er* and control conditions. This is contrary to our hypothesis. Furthermore, previous studies (Stowe, 1986; Crain & Fodor, 1985) have found that the WH-mean is higher than the control mean. Since our data does not exhibit this, our data fails to replicate the findings of these studies.

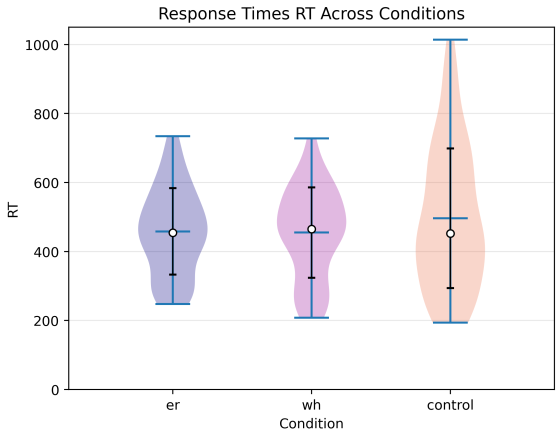

Figure 3.3: Reaction Time Across Conditions Blue line: mean, Dot: median, Error bars: SD

To examine a possible reason for this inconsistency, we can look at Fig. 3.4. This graph shows the RT of the different participants, ordered by means, from low to high. It shows us that performance varied greatly between participants. In fact, we see that the lowest two participants are almost completely responsible for the smaller peak we see in 3.1 and 3.2. This suggests that the bimodal distribution of our dataset is a result of one of two

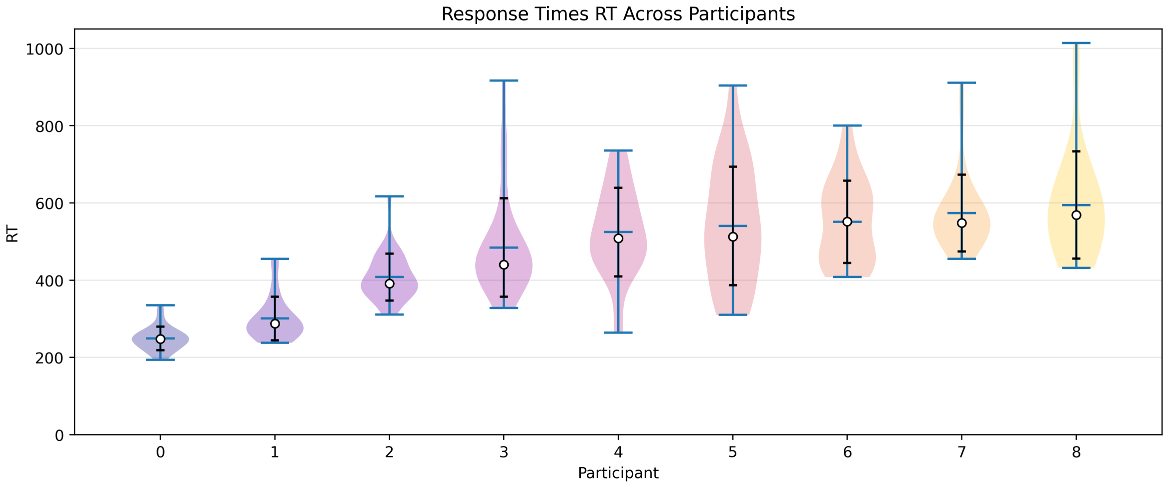

Figure 3.4: Reaction Time Across Participants, Ordered from Lowest to Highest Mean Blue line: mean, Dot: median, Error bars: SD

things:

1. The low sample size of nine resulted in the misrepresentation of data through human variation. In other words, it was coincidence that there was no participants to fill the gap between the two RT peaks.

2. The two ranges of RT stem from a difference in approach. Some participants prioritised speed over comprehension, others vice versa. This is a documented psychological effect (Heitz, 2014) and it is plausible that it affects these results as well.

We will discuss these possibilities and their implications more in Section 4.

3.3 Statistical testing

While the descriptive examination of our dataset already shows us that the data likely does not agree with our hypothesis, we will perform a series of statistical tests to formalise the results.

First, we test whether the mean of RT in WH sentences is greater than the mean of RT in control sentences. This effect is well-documented and if our data is reliable, it should be correct. To do this, we perform a one-tailed Wilcoxon Rank Sum test (U=-0.640, p=0.739). We perform this test for multiple reasons. Firstly, it is a non-parametric test, so it does not assume normality. Therefore, our bimodal data is compatible with this test. Secondly, the test compares the independence of two distributions in a single direction, which is applicable to the effect we want to test. We fail to reject the null hypothesis that the mean WH RT equals the mean control RT. This shows that our data is incongruent with the established theory.

Next, we test the main assumption of our hypothesis. Namely, we test whether the mean of RT in *er* sentences is greater than the mean of RT in control sentences. Again, we perform a one-tailed Wilcoxon Rank Sum test (U=-0.502, p=0.692). We fail to reject the null hypothesis that the mean *er* RT equals the mean control RT.

4 Discussion

In this project, we examined the processing effect of *er*, specifically the LDD of er_P. To do this, we designed a moving window self-paced reading experiment. This experiment collected the response times of participants at the moment an LDD is resolved. We hypothesised that there would be a significant slowdown in processing as a result of the LDD of *er*. However, as we have seen, the results do not support this. In fact, the results show no significant difference between any of the experimental conditions.

This means that our results do not support our hypothesis. However, we do have to consider that our experimental design may be flawed. This is especially important because our results do not replicate the findings of earlier research. Stowe (1986) and Crain & Fodor (1985) examined (among other things) the processing effects of LDD's in WH-phrases. Both concluded that a WH-phrase with an LDD increased processing load compared to a sentence without LDD. Furthermore, for statistical analyses these researchers used ANOVA's and one-tailed t-tests respectively. From this, we can infer that their results were normally distributed. Our data shows no difference in processing load between WH and control conditions, and has a bimodal distribution. This difference in result can be explained by one of two hypotheticals:

The first option is that the results and conclusions of the previous research is incorrect, and we have found an effect they failed to capture. The second option is that our experimental design is unreliable and the contradicting results are a consequence of that. We will now discuss these two options in this order.

If our experiment was sufficient and our data are reliable representations of this effect, that means there is no processing load increase when resolving an LDD of er_P. This could be explained by the fact that sentences with er_X are read faster than sentences without (Grondelaers, 2020). There has been no study on the effect of er_P on overall reading speed, but since people do not actively discriminate between the different forms of *er* while reading, it is not unlikely that a similar effect applies.

To examine the second option (that our experiment design is flawed) we will now discuss the limitations of our experiment.

4.1 Limitations

One effect that may reduce the reliability of our results is the speed-accuracy trade-off (SATO) (Heitz, 2014). This effect is caused by a difference in priority for participants of a timed experiment. Some participants prioritise completing a task quickly, while others prioritise completing it accurately. This was not taken into account during experiment and instruction design, and as such it is likely that there was a difference in approach between our participants.

This is especially supported by the fact that our data are bimodally distributed. The two peaks that form are from participants with two different approaches. This is also visible in Fig. 3.4, in which we see that the first two participants' RT's are located almost entirely within the 200-400ms range, while the other participants peak – and are mostly located – in the 400-600ms range. These two categories line up with the two peaks visible in the graphs of Fig. 3.2.

While it is possible that these differences occur naturally, it is important to note that prior to the experiment, participants were instructed to "move on to the next word once they had processed it", with little emphasis put on the speed of their response. This phrasing was inadequate and did not provide proper instruction for the experiment.

Another possible limitation of our experiment is related to the findings of Grondelaers (2020), which detail that sentences that contain er_X are read faster overall than sentences without er_X. Our experiment tests er_P, but since *er* can have multiple forms at once, it is possible to include an er_{XP} in our stimuli. However, our stimulus design only allows *er* to appear after the subject. Er_X can only appear in one of two places: at the start of the sentence, or after the primary verb (under certain conditions). Therefore er_X cannot appear in our stimuli, and as such this effect is not a confounder in our current experiment.

Finally, it is possible there was a problem of sample size. Our experiment tested on only nine participants, and as such the noise present in our dataset was a large component of the variation in our results. Therefore, if the effect we were looking for was small enough, it is possible that we could not have detected it at all with the current data.

4.2 Future Works

Due to the limitations, no reliable conclusions can be drawn about the effect of LDD's with er_p on processing speed. Therefore, the results of this paper have little *direct* practical use. However, the methods used in this experiment are functional as a baseline from which future experiments can be built. We recommend taking the following measures to improve the reliability of future research:

After completing the experiment multiple participants reported they had found the three conditions, though none had found exactly what was being measured. This may have influenced the results, though we cannot examine to what degree. There are two ways to mitigate this which we did not account for in our design. Firstly, more fillers can be added to reduce the proportion of trials that adhere to the strict stimulus design rules. Secondly, inquiring about the grammaticality of the sentence may only direct the participants' attention *more* to the structure of the sentence, instead of leading it away.

Relating to sample size, we did not perform a power analysis to calculate a necessary sample size for this experiment, which made it more difficult to determine whether sample size was a relevant factor in our null results. Since the scale of the effect can be reasoned from previous research, we recommend performing an a priori power analysis to determine the desired sample size.

Most importantly, it is likely that our instruction design led to a large difference in SATO between participants, which introduced a confounding variable into our results. We recommend mitigating this by prioritising speed in the instruction of participants.

References

Baayen, R. H., Milin, P., & Ramscar, M. (2016, November). Frequency in lexical processing. *Aphasiology*, *30*(11), 1174–1220. Retrieved 2024-07-21, from https://www.tandfonline.com/doi/full/10.1080/02687038.2016.1147767 doi: 10.1080/02687038.2016.1147767

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748–769. Retrieved 2024-03-08, from https://doi.apa.org/doi/10.1037/0278-7393.28.4.748 doi: 10.1037/0278-7393.28.4.748

Brysbaert, M., Mandera, P., & Keuleers, E. (2018, February). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, *27*(1), 45–50. Retrieved 2024-07-21, from http://journals.sagepub.com/doi/10.1177/0963721417727521 doi: 10.1177/0963721417727521

Crain, S., & Fodor, J. D. (1985, May). How can grammars help parsers? In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural Language Parsing* (1st ed., pp. 94–128). Cambridge University Press. Retrieved 2024-07-18, from https://www.cambridge.org/core/product/identifier/CBO9780511597855A040/type/book_part doi: 10.1017/CBO9780511597855.004

De Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023, May). jsPsych: Enabling an Open-Source CollaborativeEcosystem of Behavioral Experiments. *Journal of Open Source Software*, *8*(85), 5351. Retrieved 2024-06-04, from https://joss.theoj.org/papers/10.21105/joss.05351 doi: 10.21105/joss.05351

Donaldson, B. C. (2008). Dutch: a comprehensive grammar | Ch. 15: Er. In *Dutch: a comprehensive grammar* (2. ed ed., pp. 366–373). London: Routledge.

Eye-Tracking with Text. (2014). In B. VanPatten, J. Jegerski, & G. D. Keating (Eds.), *Research methods in second language psycholinguistics* (pp. 69–82). New York: Routledge.

Findelsberger, E., Hutzler, F., & Hawelka, S. (2019, July). Spill the load: Mixed evidence for a foveal load effect, reliable evidence for a spillover effect in eye-movement control during reading. *Attention, Perception, & Psychophysics*, *81*(5), 1442–1453. Retrieved 2024-06-04, from http://link.springer.com/10.3758/s13414-019-01689-5 doi: 10.3758/s13414-019-01689-5

Grondelaers, S. (2020, October). Verwacht het onverwachte: Over hoe Nederlanders er anders gebruiken dan Vlamingen. *Nederlandse Taalkunde*, *25*(2), 205–211. Retrieved 2024-03-08, from https://www.aup-online.com/content/journals/10.5117/NEDTAA2020.2-3.006.GRON doi: 10.5117/NEDTAA2020.2-3.006.GRON

Grondelaers, S., Brysbaert, M., Speelman, D., & Geeraerts, D. (2002). Er als accessibility marker: on-en offline evidentie voor een procedurele duiding van presentatieve zinnen. *Gramma/Ttt (Dordrecht)*, *9*, 1–22.

Grondelaers, S., Speelman, D., Drieghe, D., Brysbaert, M., & Geeraerts, D. (2009, February). 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. Retrieved 2024-07-22, from `https://linkinghub.elsevier.com/retrieve/pii/S0001691808001522` doi: 10.1016/j.actpsy.2008.11.003

Heitz, R. P. (2014, June). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*. Retrieved 2024-07-15, from `http://journal.frontiersin.org/article/10.3389/fnins.2014.00150/abstract` doi: 10.3389/fnins.2014.00150

Jegerski, J. (2014). Self-Paced Reading. In B. VanPatten (Ed.), *Research methods in second language psycholinguistics* (pp. 20–49). New York: Routledge.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. Retrieved 2024-07-22, from `https://doi.apa.org/doi/10.1037/0033-295X.87.4.329` doi: 10.1037/0033-295X.87.4.329

Odijk, J. E. J. M. (1993). Compositionality and syntactic generalizations. *Ph.D. Thesis, Tilburg University*.

Robinson, P. J., & Ellis, N. C. (Eds.). (2008). *Handbook of cognitive linguistics and second language acquisition*. New York: Routledge. (OCLC: ocn154308889)

Stowe, L. A. (1986, July). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245. Retrieved 2024-03-08, from `http://www.tandfonline.com/doi/abs/10.1080/01690968608407062` doi: 10.1080/01690968608407062

Vasishth, S. (2006, January). On the proper treatment of spillover in real-time reading studies: Consequences for psycholinguistic theories.

A Frequency Coefficient Ranges

Word Type	Min	Max	Outliers
Preposition	95	100	–
Noun	12	50	[1.94, 8.10, 89.77]
Adjective	25	100	[5.08]
Adverb	10	100	–
Auxiliary Verb	89	100	–
Lexical Verb	28	100	[17.48]
Numerical	98.95		–
Pronoun	75	100	–

B Stimuli condition 1: *er*

1. De politie gaat er grondig onderzoek naar doen.
2. Ze heeft er nog steeds erg veel last van gehad.
3. Oscar hoopt er ruim een jaar mee te doen.
4. Emma heeft er niet genoeg tijd voor genomen.
5. We kunnen er helaas niets meer aan doen.
6. Niels heeft er bovendien een bedrijf mee gestart.
7. Zijn vrouw moet er juist al weken op wachten.
8. Hij heeft er vast maar een deel van verteld.
9. Hij hoopt er veel nieuwe leden voor te trekken.
10. Hun kinderen zijn er twee dagen geleden mee vertrokken.
11. Danica kan er niet erg makkelijk aan trekken.
12. De meeste studenten wilden er absoluut niets mee te maken hebben.
13. Het kind kan er uiteindelijk niets van begrijpen.
14. Onderzoekers hebben er al jaren niet naar gekeken.
15. De vrienden staan er nu al uren op te wachten
16. Zij heeft er sinds gisteren niet meer aan gedacht.
17. Jan is er dag en nacht druk mee bezig geweest.
18. Mijn moeder heeft er toch een hoop geld aan overgehouden.
19. Lukas heeft er eigenlijk weinig aan gehad.
20. Hij had er anders wel een goede kans op gehad.
21. Mijn vader is er gisteren al even aan begonnen.
22. De kinderen willen er heel graag voor werken.
23. De school heeft er eindelijk een plan voor gemaakt.
24. Tim is er zaterdag opnieuw mee gekomen.
25. Ik heb er trouwens drie foto's van gemaakt.
26. Hannah heeft er een heel overdreven verhaal over geschreven.
27. De wereld heeft er helaas geen behoefte aan gehad.
28. De monteur zal er binnen drie werkdagen naar kijken.
29. Ik wilde er morgen echt mee gaan beginnen.
30. De stad heeft er afgelopen dinsdag over besloten.

C Stimuli condition 2: *WH*

1. Waar gaat de politie grondig onderzoek naar doen?
2. Waar heeft ze nog steeds erg veel last van gehad?
3. Waar hoopt Oscar ruim een jaar mee te doen?
4. Waar heeft Emma niet genoeg tijd voor genomen?
5. Waar kunnen we helaas niets meer aan doen?
6. Waar heeft Niels bovendien een bedrijf mee gestart?
7. Waar moet zijn vrouw juist al weken op wachten?
8. Waar heeft hij vast maar een deel van verteld?
9. Waar hoopt hij veel nieuwe leden voor te trekken?
10. Waar zijn hun kinderen twee dagen geleden mee vertrokken?
11. Waar kan Danica niet erg makkelijk aan trekken?
12. Waar wilden de studenten absoluut niets mee te maken hebben?
13. Waar kan het kind uiteindelijk niets van begrijpen?

14. Waar hebben onderzoekers al jaren niet naar gekeken?
15. Waar staan de vrienden nu al uren op te wachten?
16. Waar heeft zij sinds gisteren niet meer aan gedacht?
17. Waar is Jan dag en nacht druk mee bezig geweest?
18. Waar heeft mijn moeder toch een hoop geld aan overgehouden?
19. Waar heeft Lukas eigenlijk weinig aan gehad?
20. Waar had hij anders wel een goede kans op gehad?
21. Waar is mijn vader gisteren al aan begonnen?
22. Waar willen de kinderen heel graag voor werken?
23. Waar heeft de school eindelijk een plan voor gemaakt?
24. Waar is Tim zaterdag opnieuw mee gekomen?
25. Waar heb ik trouwens drie foto's van gemaakt?
26. Waar heeft Hannah een heel overdreven verhaal over geschreven?
27. Waar heeft de wereld helaas geen behoefte aan gehad?
28. Waar zal de monteur binnen drie werkdagen naar kijken?
29. Waar wilde ik morgen echt mee gaan beginnen?
30. Waar heeft de stad afgelopen dinsdag over besloten?

D Stimuli condition 3: *control*

1. De politie gaat grondig onderzoek naar de gevolgen doen.
2. Ze heeft nog steeds erg veel last van het water gehad.
3. Jan is dag en nacht druk met zijn project bezig geweest.
4. Emma heeft niet genoeg tijd voor haar onderzoek genomen.
5. We kunnen helaas niets meer aan onze resultaten doen.
6. Niels heeft bovendien met de prijs een bedrijf gestart.
7. Zijn vrouw moet juist al weken op haar resultaten wachten.
8. Hij heeft vast maar een deel van het verhaal verteld.
9. Hij hoopt veel nieuwe leden voor zijn partij te trekken.
10. Hun kinderen zijn twee dagen geleden met hun geld vertrokken.
11. Danica kan niet erg makkelijk aan de rol papier trekken.
12. De meeste studenten wilden absoluut niets met de partij te maken hebben.
13. Het kind kan uiteindelijk niets van het verhaal begrijpen.
14. Onderzoekers hebben al jaren niet naar dat probleem gekeken.
15. De vrienden staan nu al uren te wachten op de foto.
16. Zij heeft sinds gisteren niet meer aan haar vriend gedacht.
17. Jan is dag en nacht druk met zijn project bezig geweest.
18. Mijn moeder heeft toch een hoop geld aan de zaak overgehouden.
19. Lukas heeft eigenlijk weinig aan de hulp gehad.
20. Hij had anders wel een goede kans op de eretitel gehad.
21. Mijn vader is gisteren al aan het artikel begonnen.
22. De kinderen willen heel graag voor een ijsje werken.
23. De school heeft eindelijk een plan voor nieuwe studenten gemaakt.
24. Tim is zaterdag opnieuw met zijn idee gekomen.
25. Ik heb trouwens drie foto's van je huis gemaakt.
26. Hannah heeft een heel overdreven verhaal over afgelopen zondag geschreven.
27. De wereld heeft helaas geen behoefte aan onze zaak gehad.

28. De monteur zal binnen drie werkdagen naar je auto kijken.
29. Ik wilde morgen echt gaan beginnen met mijn boek.
30. De stad heeft afgelopen dinsdag over het geldprobleem besloten.

E Filler Stimuli

1. Olaf wilde daar gisteren al om gaan gevochten.
2. Ik heb er er in twee dagen al zeventien van kapotgemaakt.
3. Waar kun je mij even kort ermee helpen?
4. Waarom heeft jouw familie daar zo veel moeite mee?
5. Het nieuwe auto kan maar de helft van de afstand afleggen.
6. Mijn rekenmachine heeft veel moeite gehad mee de som.
7. De mok ligt op de groot kast in de bijkeuken.
8. Frits kan er helaas niet lang meer mee om vechten.
9. Het brood is van de hoge meeste kwaliteit graan gemaakt.
10. Waar heeft de vrouw de tijd zo snel laten gaan?
11. Hoe moeten we daar dan met omgaan?
12. Nienke is altijd al veel erg aardig tegen hem geweest.
13. Waarom willen jullie daar zo de nadruk op leggen?
14. De koelkasten blijkt nu al weken te zoemen.
15. Waar gaat Tim zo laat op de avonden heen?