Comparing linguistic theories on 'er' with what native Dutch speakers consider acceptable

Bachelor's Project Thesis

Ella Karina Collins, s3327469, e.k.collins@student.rug.nl
Supervisor: Dr. S.M. Jones, s.m.jones@rug.nl
Department of Artificial Intelligence, University of Groningen
9747 AG Groningen, The Netherlands

Abstract: The function of the Dutch pronoun *er* is complicated. Understanding the grammatical intricacies of words like *er* provides necessary information to accurately model language processing. Linguistic theories argue for four functions: existential, locative, quantitative and pronominal. According to these theories, the placement of the word *er* in a sentence is dictated by the function it possesses. Additionally, a single occurrence of *er* can assume multiple functions in certain cases. This paper aims to provide more information on the grammatical intricacies of *er* by looking at whether native Dutch speakers agree with linguistic theories of *er* using an acceptability judgement task. The focus was on the functions *er* can provide while present in the prefield. Whether other instances of *er* can occur in the sentence, either implicitly or explicitly, was also investigated. The experiment was a 4-option Likert scale task, with 43 participants taking part in the survey. Results show that existential *er* is the most acceptable function in the prefield according to native speakers. Additionally, er_{XL} and er_{XP} in the prefield are significantly more acceptable than er_{XQ}. However, er_{XQ} in the prefield is considered far more acceptable than initially expected. Finally, er_X in the prefield with er_Q in the midfield is significantly more acceptable than er_X in the prefield with er_L or er_P in the midfield. However, er_L and er_P in the midfield with er_X in the prefield do not behave as expected, with little consensus on the acceptability. The results suggest linguistic theories do not accurately describe what native Dutch speakers consider acceptable, instead there is a gradience for acceptability.

1 Introduction

The Dutch pronoun *er* has four possible functions: existential er_X, locative er_L, pronominal er_P and quantitative er_Q.

Existential er_X translates to *there* and is closely related to the English existential *there*. According to Grondelaers, Speelman, Drieghe, Brysbaert, and Geeraerts (2009), it has evolved from the locative adverb "daar", *there* in English, and can be used to introduce new unexpected information in a sentence. Er_X can occur in a subjectless passive sentence or in passive sentences with a subordinate clause (Odijk, 1993). It can also occur in sentences with indefinite subjects (Donaldson, 2008). This is the case in (1), where "een man" is the indefinite subject.

(1)   *Er_X loopt een man op straat.*
      ER   walks a   man on street
      "There is a man walking in the street."

Locative er_L also translates to *there* and is closely related to the English locational adverb *there*. In Dutch, er_L can replace an adverbial locative phrase, like "daar" (Odijk, 1993; Donaldson, 2008; Webelhuth and Bonami, 2019). In sentence (2), locative er_L is referring to the location the person has been living.

(2)   *Hij heeft er_L tien jaar gewoond.*
      He has   ER ten   years lived
      "He lived there for ten years."

Pronominal er_P serves as a prepositional object for the sentence, replacing a pronoun when specified by the preposition (Odijk, 1993). According

to Webelhuth and Bonami (2019), er_P is an obligatory compliment of the Dutch preposition *op*. It can replace the pronoun "het", *it* or *them* (Odijk, 1993; Donaldson, 2008). Er_P can either be adjacent to or separate from its selector in a sentence. In (3), er_P is separate from its selector "mee".

(3)   Ik heb er_P het brood mee gesneden.
      I   have ER the bread with cut

      "I cut the bread with it."

Quantitative er_Q appears in sentences with headless noun phrases where the variable is a count noun (Odijk, 1993). It performs a similar function as the partitive element *en* in French or *ne* in Italian (Webelhuth and Bonami, 2019). According to Donaldson (2008), er_Q is used as a complement for numerals and quantity adverbs. In (4), er_Q is the compliment of the number "vijf".

(4)   Jan heeft er_Q vijf.
      Jan has   ER five

      "Jan has five of them."

Additionally, a single occurrence of *er* can assume multiple functions in a sentence. In (5), *er* has both a quantitative and a locative function, where one function is implicit. In this case the locative function is implicit.

(5)   Hij kocht er_{QL} slechts twee.
      He bought ER   only   two

      "He bought only two (books) there."

Linguistic theories on *er* argue that the placement of *er* in the sentence is specific to which function it assumes. Dutch sentences include a prefield and a midfield (Webelhuth and Bonami, 2019). The general structure of a Dutch main clause is shown in (6a), the prefield is followed by the inflected verb, then the midfield and finally any additional verbs necessary. In (6b) "ik" is in the prefield, "kijk" is the inflected verb, and "even op de computer" is in the midfield. In this case, there are no additional verbs necessary.

(6)   a.   prefield — inflected verb — midfield — (additional verbs)

      b.   Ik kijk even   op de   computer.
           I   look for a bit on the computer

           "I take a look at the computer."

In the prefield, existential er_X is the only function allowed to appear explicitly (Odijk, 1993; Neeleman and Van de Koot, 2006; Webelhuth and Bonami, 2019). While er_X is in the prefield it can assume the locative or pronominal functions. However, it cannot assume the quantitative function (Odijk, 1993; Neeleman and Van de Koot, 2006; Webelhuth and Bonami, 2019).

In the midfield, any function of *er* is possible. However, only one *er* can appear explicitly, which can implicitly assume any other necessary functions (Neeleman and Van de Koot, 2006; Webelhuth and Bonami, 2019). Grondelaers et al. (2009) further suggests that er_X in the midfield provides a processing advantage for new information.

This paper will be comparing these linguistic theories with the opinions of native Dutch speakers. Therefore, the research question being investigated here is "How closely do linguistic theories of *er* correspond with what native Dutch speaker consider acceptable?" This question can be broken down into multiple sub-questions looking at specific placements of *er* in the sentence. This paper will focus on the constraints when *er* is present in the prefield. This involves looking at the possible functions *er* can have while occupying the prefield position and whether other instances of *er* can occur in the sentence.

Therefore, the first sub-question that will be considered is "When *er* in the prefield takes on one function, which of the functions are allowed?" As discussed previously, linguistic theories state only existential er_X can appear explicitly in the prefield (Odijk, 1993; Neeleman and Van de Koot, 2006; Webelhuth and Bonami, 2019). Therefore, it can be hypothesised that, while present in the prefield, locative er_L, quantitative er_Q and pronominal er_P will not be accepted by native Dutch speakers.

Once the possible functions for *er* in the prefield have been established, then the presence of multiple occurrences of *er* in a sentence can be investigated. This includes looking at whether er_X in the prefield must assume the functions of er_L, er_P or er_Q, and whether er_L, er_P or er_Q can appear in other parts of the sentence. Therefore, the second sub-question being investigated is "Can er_X in the prefield assume any other *er* functions?" (This is termed implicit *er*.) According to linguistic theories, with er_X in the prefield implicit er_L or er_P are possible (Neeleman and Van de Koot, 2006; Webel-

huth and Bonami, 2019). Therefore, the hypothesis here is that native Dutch speakers will find er_{XP} or er_{XL} in the prefield acceptable, but not er_{XQ}.

The third sub-question to consider is "With er_X in the prefield, can other explicit ers appear in the midfield?" Linguistic theories argue only an explicit er_Q can appear in the midfield when er_X occupies the prefield (Odijk, 1993; Neeleman and Van de Koot, 2006; Webelhuth and Bonami, 2019). Therefore, it can be hypothesised that native Dutch speakers will not find er_P or er_L in the midfield with er_X in the prefield acceptable, while er_Q in the midfield will be found acceptable.

An acceptability judgement task will be carried out to gather data on the opinions of native Dutch speakers. Testing the discussed hypotheses using these opinions proves difficult. It is hard to establish a level for complete acceptability and unacceptability to compare with the opinions. In this paper, filler sentences are used as the standard for completely acceptable and completely unacceptable. The conditions will be tested against these standards to see whether they are considered acceptable. The filler sentences and conditions are explained in the methods section below.

2 Methods

2.1 Acceptability Judgement Tasks

Acceptability judgement tasks ask participants to provide an opinion on the acceptability of certain sentences. There are many different tasks that can be conducted for an acceptability judgement. This experiment will use a Likert scale task. A Likert scale task provides participants with a numerical scale with a chosen amount of options, consistent across all questions. The endpoints of the scale are fully acceptable or fully unacceptable. A Likert scale benefits from being both intuitive and numerical, causing it to be effective for answering questions on the "size of a difference" (Sprouse and Almeida, 2013).

According to Croasmun and Ostrom (2011), scales with an odd number of options are more frequently used than even-numbered scales. However, even-numbered scales are perfectly acceptable. Odd-numbered Likert scales provide participants with a neutral choice giving them the ability to not have an opinion, whereas even-numbered Likert scales force participants to choose a side (Brown, 2000).

In this experiment forcing a choice is desirable. Therefore, a four choice scale was created. The scale used is as follows:

1. completely unacceptable
2. probably unacceptable
3. probably acceptable
4. completely acceptable

This scale is translated into Dutch for the experiment to reduce confusion for participants. The Dutch translation is as follows:

1. volledig onacceptabel
2. waarschijnlijk onacceptabel
3. waarschijnlijk acceptabel
4. volledig acceptabel

2.2 Stimuli

The stimuli sentences are first separated into two categories: target sentences and filler sentences. In total, there are 80 stimuli sentences, 40 target sentences and 40 filler sentences.

2.2.1 Target Sentences

The 40 target sentences are ordered into three categories: er in the prefield with a single function, er_X in the prefield with an implicit function, and er_X in the prefield with an explicit er occurring in the midfield. Four sentences were created for each of these conditions. These sentences were checked by a native Dutch speaker to ensure they were constructed correctly.

The first category, er in the prefield with a single function, consists of four conditions: er_X, er_L, er_P and er_Q. An example of each condition is shown in (7).

(7)　a.　*Er$_X$ is veel gebeurd hier.*
　　　　ER　is a lot happened here

　　　　"There is a lot happening here."

　　b.　* *Er$_L$ woont ze al tien jaar.*
　　　　ER　lived she already ten year

　　　　(intended) "She has lived there for ten years."

3

c. * *Er$_P$ kijk ik vaak naar.*
ER look I often to

(intended) "I often watch it."

d. * *Er$_Q$ heeft Sara drie gemaakt.*
ER has Sara three made

(intended) "Sara has made three of them."

The second category, er_X in the prefield with an implicit function, consists of three conditions: implicit locative er_{XL}, implicit pronominal er_{XP} and implicit quantitative er_{XQ}. An example sentence for each condition is shown (8).

(8) a. *Ken je Upsilon? Er$_{XL}$ wonen*
Know you Upsilon? ER live
veel studenten.
many students

"Do you know Upsilon? Many students live there."

b. *Er$_{XP}$ keken veel mensen naar.*
ER looked many people to

"There were many people watching it."

c. * *Er$_{XQ}$ lagen drie op tafel.*
ER lay three on table

(intended) "There were three of them on the table."

Er_X in the prefield with an explicit *er* occurring in the midfield also consists of three conditions: er_X in the prefield with er_L in the midfield ($er_X er_L$), er_X in the prefield with er_P in the midfield ($er_X er_P$) and er_X in the prefield with er_Q in the midfield ($er_X er_Q$). An example of each condition can be seen in (9).

(9) a. * *Ken je Middelstum? Er$_X$*
Know you Middelstum? ER
wonen er$_L$ veel boeren.
live ER many farmers

(intended) "Do you Middelstum? There are many farmers living there."

b. * *Er$_X$ denken er$_P$ mensen aan.*
ER think ER people on

(intended) "There are many people thinking about it."

c. *Er$_X$ zijn er$_Q$ twee gestolen.*
ER were ER two stolen

"There were two of them stolen."

For the condition er_{XL}, where existential er_X has an implicit locative function, finding Dutch verbs that required a location was difficult. Therefore, to establish an implicit location an contextual sentence was included, (10) shows an example of such a sentence. These contextual sentences were also added to the condition $er_X er_L$, where existential er_X is in the prefield and locative er_L is in the midfield, such as sentence (11). This was done to ensure that locative er_L was treated the same in the two different categories. Additionally, certain sentences were paired between categories, including (10) and (11).

(10) *Ken je de Korreweg? Er$_{XL}$ gebeuren*
Know you the Korreweg? ER happen
ongelukken.
accidents

"Do you know the Korreweg? Accidents happen there."

(11) * *Ken je de Korreweg? Er$_X$*
Know you the Korreweg? ER
gebeuren er$_L$ ongelukken.
happen ER accidents

(intended) "Do you know the Korreweg? Accidents happen there."

2.2.2 Filler Sentences

The filler sentences do not include the word *er*. They are further categorised as bad fillers and good fillers, with 20 sentences each. It is important to include these control sentences to reduce and detect non-cooperative participants (Häussler and Juzek, 2017). These are also included to ensure participants make use of the extreme options of the scale.

The bad filler sentences were created by first constructing sentences in Hindi and translating them word for word into Dutch, an example is shown in (12). This provided sentences with a completely different structure that should result in the "completely unacceptable" judgement. On the other hand, the good filler sentences were taken from a beginners Dutch text book *Nederlands in gang* (2010), an example is shown in (13).

(12) * *Twee de honden het deel weg*
Two the dogs it part away

(intended) "Two dogs ran away."

(13) *Kunt u   mij helpen? Ik zoek   een*
      Can you me help?   I   search a
      spijkerbroek.
      jeans

      Can you help me? I am looking for jeans.

All sentences used as stimuli in the experiment are listed in Appendix A.

2.3 The Experiment

The experiment was created as an online survey on Qualtrics, an online survey tool. Participants were first shown a consent page stating if they choose to continue they consent to participating in the experiment. Participants are then asked to enter their age, place of residence and place of origin, after which the participants are provided with instructions on how to complete the survey. The instructions, given in Dutch, asked them to judge each sentence by how acceptable they found it and choose the option that best reflects their judgement.

The 80 sentences are randomized and shown one at a time on the screen. The participants must choose one of the four options before being able to proceed to the next sentence. It is not possible to return to a previously answered question. Additionally, after choosing a judgement option, participants had to click 'continue' to continue to the next sentence. This forces them to move their mouse to a different location, which increases participant reliability by preventing repetitive clicking to go through the survey (Häussler and Juzek, 2017).

2.3.1 Participants

Participants were recruited online. In total 43 native Dutch speakers took part in the experiment. Place of origin showed participants grew up in the Netherlands. The average age was 21.5 years. No data on gender or bilingualism was collected. The majority of participants were university students. Individual links were created for each participant. This provided the ability to ensure that each participant completed the survey. Each participant was compensated with 5 euros for participation.

3 Results

3.1 Data

Likert scale tasks produce ordinal data. This type of data is non-continuous ranked categorical data. Therefore, it is necessary to compare the medians (Mdn) and modes of the conditions, instead of the means. Furthermore, the experiment consisted of one independent variable, the function *er* possesses, with multiple levels. A Kruskal-Wallis test is used in cases with one independent variable that has two or more levels, where the data is ordinal or interval (Bruin, 2011). Therefore, this test will be used to identify significant differences. For this test, the test statistic is given by H, the Kruskal Wallis chi-square statistic.

3.2 Filler Sentences

The results for the filler sentences help determine participant reliability. If a participant judged bad filler sentences with a high acceptability or good filler sentences with a low acceptability, the participant would be considered unreliable and their data would be excluded from analysis.
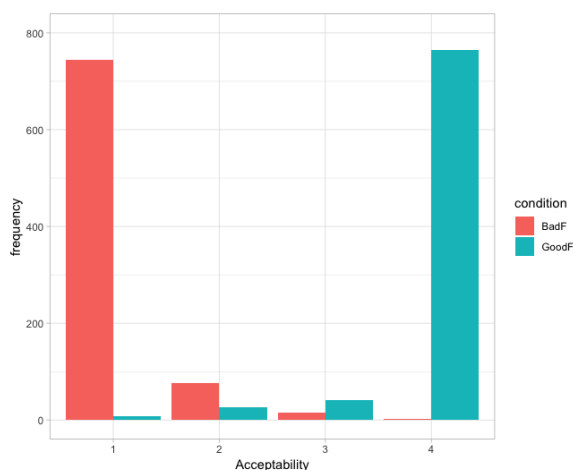

Figure 3.1: Filler sentences frequency graph

Figure 3.1 shows the frequency for good and bad filler sentences. The bad filler sentences (*BadF*) are shown in red, while the good filler sentences (*GoodF*) are shown in blue. The data was aggregated to retrieve the median values. The good filler

sentences have a median value of 4, completely acceptable. The bad filler sentences have a median value of 1, completely unacceptable. These median values are as expected, suggesting that participants overall found bad filler sentences unacceptable and good filler sentences acceptable.

However, individual participant data must be examined to ensure data reliability. One participant consistently ranked bad filler sentences with values of 3 or 4, probably acceptable or completely acceptable. These sentences should not be considered acceptable, suggesting that this participant is unreliable. Therefore, the participant is considered an outlier and their data was removed for the rest of the analysis.

3.3 Target Sentences

Firstly, looking at the raw data in the form of frequency graphs for the three categories of conditions will provide an idea of the distributions of the data.
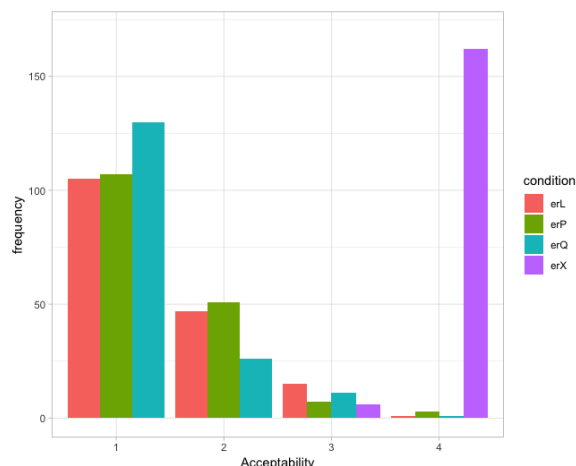
Figure 3.2: Frequency graph for *er* in the prefield with a single function

Figure 3.2 shows the frequency for the condition where *er* with a single function is in the prefield. This graph shows er_X in the prefield has the highest acceptability, with most judgements of 4, 'completely acceptable, and some judgements of 3, 'probably acceptable'. The other three conditions have judgements across all acceptabilities, with a majority for 1, 'completely unacceptable'.

Figure 3.3 shows the frequencies for existential er_X in the prefield with an implicit function. This
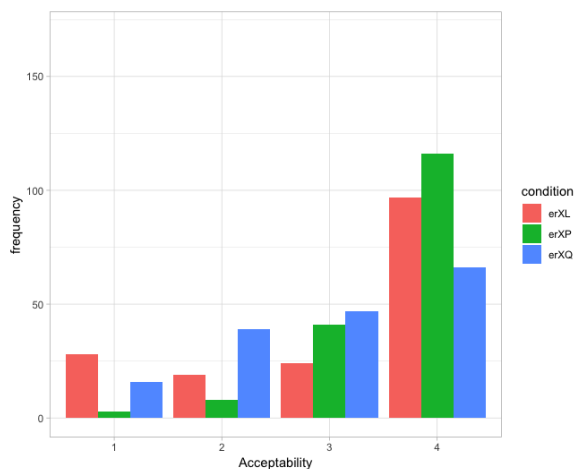
Figure 3.3: Frequency graph for existential *er* in the prefield with an implicit function

graph shows little differences in acceptability between the conditions. It seems existential *er* in the prefield with an implicit quantitative function, er_{XQ}, is less acceptable than existential *er* in the prefield with an implicit locative function, er_{XL} or an implicit pronominal function, er_{XP}. Additionally, the graph shows that er_{XP} is the most acceptable of three conditions.
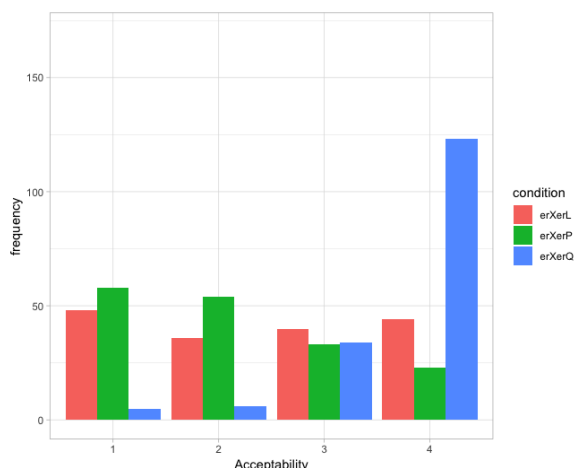
Figure 3.4: Frequency graph for existential *er* in the prefield with another explicit *er* in the midfield

Figure 3.4 shows the frequencies for existential er_X in the prefield with an explicit *er* occurring

in the midfield. This graph shows that er_X in the prefield with er_Q in the midfield ($er_X er_Q$), is the most acceptable of the three conditions. However, there seems to be no consensus on the acceptability for the other two conditions. For er_X in the prefield and er_P in the midfield ($er_X er_P$), the distribution increases towards unacceptable, whereas the acceptability for er_X in the prefield and er_L in the midfield ($er_X er_L$), seems to have no pattern.

After looking at the graphs of the raw data, the medians for each condition were calculated. These values can be seen in figure 3.5. Also, the mode values were found for each condition, shown in figure 3.6. The median and mode values can range from completely unacceptable with a value of 1 to completely acceptable with a value of 4.
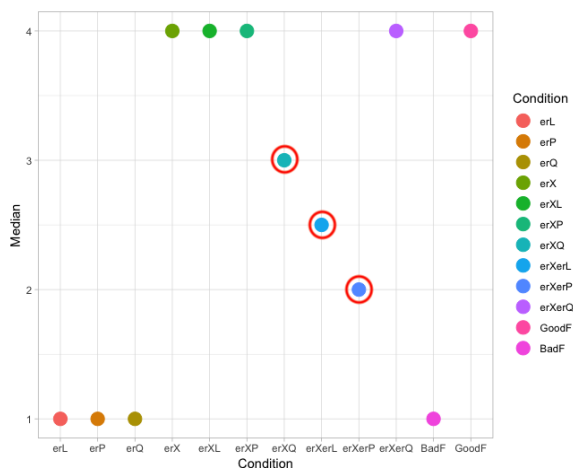

Figure 3.6: Mode values for all conditions
Unpredicted values circled in red

Figure 3.5: Median values for all conditions
Unpredicted values circled in red

3.3.1 Results in Table 3.1

Each condition has a theoretical expected acceptability. Using the data from the filler sentences, these expectations can be tested. Filler sentence data is used here because it shows natural acceptability for acceptable sentences, good filler sentences ($GoodF$), and unacceptable sentences, bad filler sentences ($BadF$). Table 3.1 shows the Kruskal-Wallis tests for significance between the expectations and the conditions. A significant difference between a condition and its expectation means that the condition does not behave as predicted. On
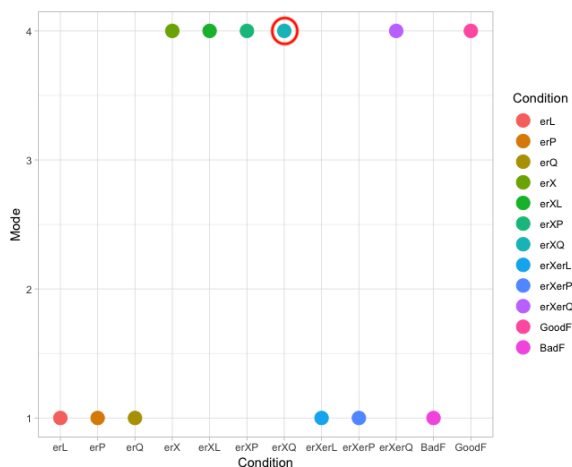
the other hand, no significant difference means the condition behaves as predicted.

conditions	H	df	p-value
er_X vs. $GoodF$	5.9018	1	0.01513
er_L vs. $BadF$	73.622	1	$< 2.2e^{-16}$
er_P vs. $BadF$	66.078	1	$4.335e^{-16}$
er_Q vs. $BadF$	16.43	1	$5.048e^{-5}$
er_{XL} vs. $GoodF$	132.25	1	$< 2.2e^{-16}$
er_{XP} vs. $GoodF$	56.745	1	$4.961e^{-14}$
er_{XQ} vs. $BadF$	539.57	1	$< 2.2e^{-16}$
$er_X er_L$ vs. $BadF$	342.65	1	$< 2.2e^{-16}$
$er_X er_P$ vs. $BadF$	274.17	1	$< 2.2e^{-16}$
$er_X er_Q$ vs. $GoodF$	39.49	1	$3.297e^{-10}$

Table 3.1: Kruskal-Wallis tests for target conditions against theoretical expectations.
H is the Kruskal-Wallis chi-squared test statistic and df is the degrees of freedom.

The first category of conditions, er in the prefield with a single function, are represented by the first four points in figures 3.5 and 3.6. Here it seems er_X in the prefield ($Mdn = 4, mode = 4$) is considered acceptable, while er_L ($Mdn = 1, mode = 1$), er_P ($Mdn = 1, mode = 1$) and er_Q ($Mdn = 1, mode = 1$) in the prefield are not considered acceptable. A Kruskal-Wallis test for significance shows a significant difference between er_X in the prefield and good filler sentences ($GoodF$). Testing er_L, er_P and er_Q in the prefield against bad filler

sentences ($BadF$) shows that there is a significant difference in acceptability. However, differences in the H statistic can be seen. Er_L against $BadF$ has the highest H value, followed by er_P against $BadF$ then er_Q against $BadF$ and finally er_X against $GoodF$.

The second category of conditions is er_X in the prefield with an implicit locative, pronominal or quantitative function. The graphs suggest that er_{XL} ($Mdn = 4, mode = 4$), er_{XP} ($Mdn = 4, mode = 4$) and er_{XQ} ($Mdn = 3, mode = 4$) are all considered acceptable to a degree. A Kruskal-Wallis test shows that there is a significant difference between good filler sentences and er_{XL}, as well as er_{XP}. Also, there is a significant difference between er_{XQ} and bad filler sentences. Here the H statistic for er_{XQ} against $BadF$ is much high than the other two conditions. However, er_{XL} against $GoodF$ also has a high value, as does er_{XQ} against $BadF$.

The third category of conditions is er_X in the prefield with an explicit er occurring in the midfield, shown by the last three points in figures 3.5 and 3.6. Here it seems $er_X er_Q$ ($Mdn = 4, mode = 4$) is considered acceptable and $er_X er_P$ ($Mdn = 2, mode = 1$) is considered unacceptable, but $er_X er L$ ($Mdn = 2.5, mode = 1$) seems undecided. A Kruskal-Wallis test shows that there is a significant difference in acceptability between bad filler sentences and $er_X er_L$ as well as $er_X er_P$. Also, $er_X er_Q$ is significantly different from good filler sentences for acceptability. Here, $er_X er_L$ against $BadF$ has the highest H value, followed by $er_X er_P$ against $BadF$. $Er_X er_Q$ against $GoodF$ has a comparatively low H value to the other two conditions.

3.3.2 Results in Table 3.2

Since all conditions are significantly different from their theoretical expectancy, none of the conditions behave as predicted. Therefore, the categories of conditions should be looked at separately. The Kruskal-Wallis tests for significance for between pairs of conditions are shown in table 3.2.

For er in the prefield with a single function, the function er possesses significantly affects how acceptable native Dutch speakers find a sentence. Unpacking this result will help see specifically where the significance lies. Separately comparing er_X in the prefield with er_L, er_P and er_Q show that er_X is

conditions	H	df	p-value
prefield	438.98	3	$< 2.2e^{-16}$
er_X vs. er_L	289.29	1	$< 2.2e^{-16}$
er_X vs. er_P	285.33	1	$< 2.2e^{-16}$
er_X vs. er_Q	298.55	1	$< 2.2e^{-16}$
er_L vs. er_P	0.18075	1	0.6707
er_P vs. er_Q	6.1991	1	0.01278
er_Q vs. er_L	7.9862	1	0.004713
implicit	36.056	2	$1.481e^{-8}$
er_{XL} vs. er_{XP}	10.798	1	0.001016
er_{XP} vs. er_{XQ}	39.61	1	$3.102e^{-10}$
er_{XQ} vs. er_{XL}	4.5252	1	0.0334
explicit	156.82	2	$< 2.2e^{-16}$
$er_X er_L$ vs. $er_X er_P$	7.5843	1	0.005888
$er_X er_P$ vs. $er_X er_Q$	145.11	1	$< 2.2e^{-16}$
$er_X er_Q$ vs. $er_X er_L$	91.439	1	$< 2.2e^{-16}$

Table 3.2: Kruskal-Wallis tests for categories and between target conditions.
H is the Kruskal-Wallis chi-squared test statistic and df is the degrees of freedom.

significantly more acceptable in the prefield. Additionally, there is no significant difference in acceptability between er_L and er_P in the prefield. Finally, separately comparing er_Q in the prefield to er_L and er_P shows that er_Q is significantly the least acceptable function to occur in the prefield.

For er_X in the prefield with an implicit function, a Kruskal-Wallis test shows that there is a significant difference in acceptability. To understand where this significance lies, separate tests between the three conditions are conducted. These tests show that er_{XP} is significantly more acceptable than er_{XL} and er_{XQ}. Additionally, er_{XL} is significantly more acceptable than er_{XQ}, showing that er_{XQ} is the least acceptable condition of the three.

Finally, for er_X in the prefield with an explicit er occurring in the midfield, the function er possesses in the midfield significantly affects the acceptability of the sentence. Looking into the separate conditions shows that $er_X er_Q$ is significantly more acceptable than $er_X er_L$, as well as $er_X er_P$. Additionally, there is a significant difference in acceptability between $er_X er_L$ and $er_X er_P$, with $er_X er_P$ being the less acceptable condition.

4 Discussion

The results show that all conditions are significantly different from their respective theoretical expectancy. This suggests that no condition behaved as originally predicted. To begin with, finding the correct statistical test to detect significance was challenging. The results found bring into question whether the correct comparisons were made with the Kruskal-Wallis test. The target sentences were compared to either good or bad filler sentences, depending on their predicted acceptability. The good and bad filler sentences were assumed to be naturally grammatical and ungrammatical. However, this may not be the case. The bad filler sentences were constructed with a different language's sentence structure. This might have led to much more ungrammatical sentences than sentences with incorrect uses of er functions. The meaning of the target sentences with a prediction of completely unacceptable may still be easily discerned, whereas the meaning of bad filler sentences may have been lost completely. Therefore, comparing the bad filler sentences to target sentences may have been incorrect and led to significance where no significance should exist.

Despite the significant difference found, the Kruskal-Wallis test still provides information on how far the conditions deviate from their theoretical predictions. The Kruskal-Wallis chi-squared test statistics, the H values, can be looked at to determine deviation. These are shown in table 3.1. A small value would suggest the conditions deviated little, while a larger value would suggest more deviation.

Comparisons between conditions in the categories were also conducted. These tests do not show whether a condition is considered completely acceptable or completely unacceptable, instead they show variance in acceptability between the conditions. The results show significant differences for acceptability between the conditions in the three categories. This suggests that there is a range of acceptability for the functions of er, a gradient.

Looking at er in the prefield with a single function, er_X is significantly different from good filler sentences ($GoodF$), with a very small H value. This suggests that the er_X condition is still similar to the good filler sentences. Looking at the raw data suggests that this significant difference shows that er_X

in the prefield is more acceptable than $GoodF$. The H value for er_Q against bad filler sentences ($BadF$) is also quite small, suggesting that er_Q is similar to bad filler sentences. On the other hand, er_P and er_L against $BadF$ have comparatively high H values, suggesting that they deviate from bad filler sentences more than er_Q. This gradience of acceptability is further proven by looking at between condition significance, see table 3.2.

Acceptability for er_X in the prefield with an implicit function also shows gradience phenomena. Here, the H value for er_{XL} against $GoodF$ is more than double the value for er_{XP} against $GoodF$, suggesting that the er_{XP} condition is closer to good filler sentences than the er_{XL} condition. Additionally, the H value for er_{XQ} against $BadF$ is very large, suggesting that er_{XQ} is vastly different in comparison to bad filler sentences. This is also evident as er_{XQ} is not significantly different from er_{XL}, see table 3.2. However, er_{XQ} is significantly different from er_{XP}, further showing the gradient where er_{XP} is the most acceptable.

Furthermore, er_X in the prefield with an explicit occurrence of er in the midfield shows gradience as well. $Er_X er_Q$ against $GoodF$ has a comparatively small H value, suggesting that this condition is relatively close to the good fillers condition. On the other hand, $er_X er_L$ and $er_X er_P$ against $BadF$ both have very large H values, suggesting that they are not found completely unacceptable. Both $er_X er_P$ and $er_X er_L$ are significantly different from $er_X er_Q$, see table 3.2. However, looking at these H values shows that $er_X er_P$ has a higher H value, suggesting that $er_X er_P$ is more vastly different than $er_X er_Q$.

Since the results suggest the presence of gradient phenomena for all of the conditions, it is important to look back at the medians and modes of the conditions, shown in figures 3.5 and 3.6, to further determine behaviour against prediction. The medians and modes provide qualitative information about the data that a yes/no test for significance does not easily communicate.

4.1 Results against prediction

Despite er_{XQ} being significantly less acceptable than er_{XP} and er_{XL}, the data does not behave as expected. The majority of the responses for the condition are acceptable. Looking into the sepa-

rate sentences for the condition may provide insight on the high acceptability. Figure 4.1 shows the frequencies per sentence. Sentence 1 ($S1$) has no judgements of 1, completely unacceptable, whereas sentence 4 ($S4$) has the highest frequency of completely unacceptable judgements. $S1$ is shown in (14) and $S4$ is shown in (15).
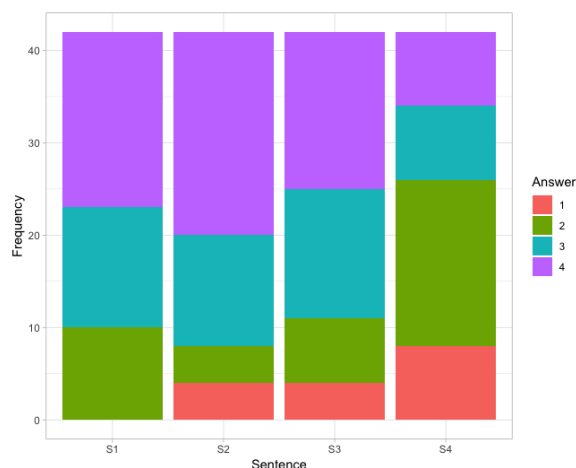

Figure 4.1: Frequency of responses per sentence for existential *er* in the prefield with an implicit quantitative function.
(1 is completely unacceptable, 2 is probably unacceptable, 3 is probably acceptable and 4 is completely acceptable)

(14)   * Er_{XQ} *zijn gisteren   twee gestolen.*
        ER   are yesterday two stolen.

     (intended) "There were two stolen yesterday."

(15)   * Er_{XQ} *heeft iemand   vier gebroken.*
        ER   has   someone four broken.

     (intended) "Someone has broken four of them."

One explanation for the lack of unacceptability for these sentences might be because the quantitative function is not necessary in these cases. For example, English sentences with headless noun phrases where the variable is a quantity do not required an equivalent pronoun to *er*. (16) shows such a sentence.

(16)   Five were made yesterday.

Another explanation comes from looking at the grammatical function the quantity plays in the sentences. In $S1$, $S2$ and $S3$, the quantity is the subject of the sentence. However, in $S4$ the quantity is the object of the sentence. As $S4$ seems to be the least acceptable of the four sentences, the grammatical role of the quantity may impact the acceptability.

Additionally, the conditions $er_X er_L$ and $er_X er_P$ do not behave as predicted. The responses for both are more positive than expected.

$Er_X er_L$ seems to be randomly distributed, shown clearly in figure 3.4. Looking at the frequencies of the separate sentences, shown in figure 4.2, also shows a rather random distribution. Here $S3$ has the highest completely unacceptable rating and the lowest completely acceptable rating. The other three sentences have similar ratings, with $S4$ having the highest acceptability. $S3$ is shown in (17) and $S4$ is shown in (18).
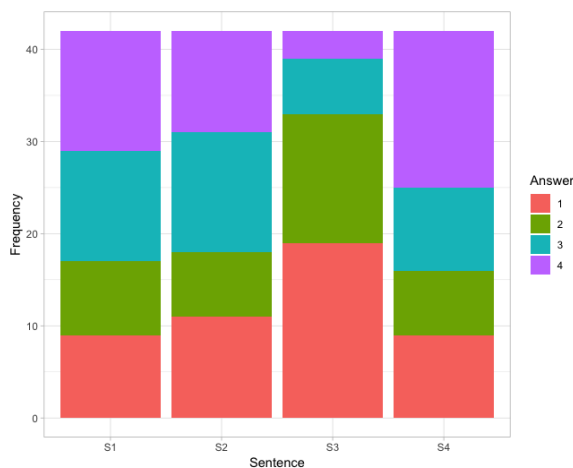

Figure 4.2: Frequency of responses per sentence for existential *er* in the prefield with locative *er* in the midfield.
(1 is completely unacceptable, 2 is probably unacceptable, 3 is probably acceptable and 4 is completely acceptable)

(17)   * *Ken   je   de   treinstation?* Er_X
        Know you the   train-station? ER

        wacht er_L *een man al dagen.*
        wait   ER a   man for days.

     (intended) "Do you know the train station? A man has been waiting there for days."

10

(18) * *Ken je Spijkenisse? Er$_X$ is er$_L$ een*
      Know you Spijkenisse? ER is ER a
      brug gebouwd.
      bridge built.

      (intended) 'Do you know Spijkenisse? A
      bridge has been built there."

Responses to $er_X er_P$ shows a slightly more unacceptable judgement, shown in figure 3.4. However, the amount of unacceptable judgements is much less than expected. Figure 4.3 shows the frequencies for the individual sentences. Here it can be seen that $S3$ has the highest completely unacceptable rating and a comparably low completely acceptable rating. $S1$ has the lowest amount of completely unacceptable and probably unacceptable ratings. $S3$ is shown in (19) and $S1$ is shown in (20).
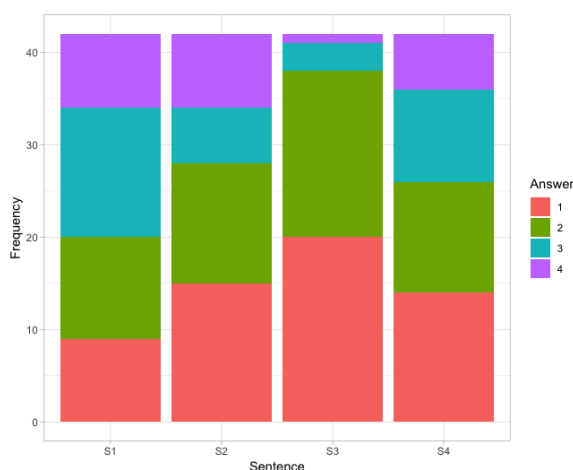

Figure 4.3: Frequency of responses per sentence for existential *er* in the prefield with pronominal *er* in the midfield.
(1 is completely unacceptable, 2 is probably unacceptable, 3 is probably acceptable and 4 is completely acceptable)

(19) * *Er$_X$ denken er$_P$ mensen aan.*
      ER think ER people on.

      (intended) "People think about it."

(20) * *Er$_X$ wordt er$_P$ morgen over*
      ER will ER tomorrow about
      gesproken.
      spoken.

      (intended) "It will be discussed tomorrow."

A potential reason for the difference in acceptability between sentences could be because certain sentences are in the passive voice. For $er_X er_L$, $S4$ is a passive sentence while the other three sentences are active. For $er_X er_P$, $S1$ is the only passive sentence in the condition. Therefore, the passive voice could cause sentences to be found more acceptable in these and other conditions.

4.2 Other issues

An additional issue specifically for the conditions with er_L can be identified. As mentioned previously, introducing a location by using er_L can be rather challenging, especially if er_L is not explicitly in the sentence. Therefore, it is possible that such sentences are not perfectly correct.

One issue that could have caused unexpected results for all conditions might be due to the small amount of stimuli sentences per condition, with only four sentences each. This could easily lead to a lack information for the conditions.

Another issue could be caused by the small size of the sample. However, it is not unusual for linguistic studies to have small samples. The sample may also be biased as only university students were tested. This can effect the generalizability of the results as not all native Dutch speakers have the same level of education.

Even though the Netherlands is a small country, there are differences in language in different parts of the country. These differences in dialects could influence the acceptability judgement of participants.

The environment in which the experiment was completed by participants can influence the quality of the results. As the survey was completed online, it was impossible to control the environment for participants. This could introduce an unpredictable variable. Also, despite the filler sentences providing results as expected, it is hard to say that all participants were fully concentrated on the experiment the whole way through. Therefore, it is impossible to assume that all the data was reliable.

Additionally, choosing the even scale forces people to have an opinion even if they have none (Brown, 2000). This could causes issues with responses to sentences where participants do not have an opinion on whether the sentence is acceptable or unacceptable. However, when it comes to grammar, most people have an opinion on whether they find

a sentence grammatical, no matter how small that opinion.

5 Conclusions

Linguistic theories provide expectations for which conditions should be found acceptable in certain positions in a sentence. Theories argue er_X is the only acceptable function to occur in the prefield. While in the prefield, er_X must assume a locative or pronominal function implicitly, but cannot assume the quantitative function. However, er_Q can appear in the midfield when er_X is in the prefield, while the other two functions are not allowed.

According to the results found in this experiment, linguistic theories do not accurately describe what native Dutch speakers consider acceptable when it comes to er. Instead of the absolute levels of acceptability that were predicted, gradient phenomena were discovered.

The results do show that existential er_X in the prefield is the most acceptable condition according to native speakers. However, locative er_L, pronominal er_P and quantitative er_Q were not considered completely unacceptable as hypothesised. Therefore, the hypothesis for er in the prefield, stating that er_L, er_P and er_Q will not be accepted by native Dutch speakers, can be rejected. There is a gradience for acceptance where er_X is the most acceptable followed by er_P, then er_L and, finally, er_Q.

Additionally, results do not show that existential er_{XL} with an implicit locative function or existential er_{XP} with an implicit pronominal function are completely acceptable. Nor do the results show that existential er_{XQ} with an implicit quantitative function is unacceptable according to native speakers. Therefore, the hypothesis that native Dutch speakers will find er_{XP} or er_{XL} in the prefield acceptable, but not er_{XQ}, must be rejected. The gradient phenomena found here shows that er_{XP} is the most acceptable, followed by er_{XL} and then er_{XQ}.

Furthermore, results show that existential er_X in the prefield with quantitative er_Q in the midfield ($er_X er_Q$) is found acceptable. However, there is little consensus on the acceptability for both existential er_X in the prefield with locative er_L in the midfield ($er_X er_L$) and existential er_X in the prefield with pronominal er_P in the midfield ($er_X er_P$). Therefore, the final hypothesis, stating that native

Dutch speakers will not find er_P or er_L in the midfield with er_X in the prefield, while er_Q in the midfield will be found acceptable, must be rejected. However, there is a gradient phenomena showing that $er_X er_Q$ is the most acceptable, followed by $er_X er_L$ and then $er_X er_P$.

The overarching research question of this paper is "how closely do linguistic theories of er correspond with what native Dutch speakers consider acceptable?" This paper has found that native speakers do not mind uses of er that linguistic theories would characterise as ungrammatical. However, the gradient phenomena found do show that grammatically correct conditions, according to linguistic theories, are more acceptable than grammatical incorrect conditions. Therefore, linguistic theories of er do not fully correspond with what native Dutch speakers consider acceptable.

5.1 Further Research

It would be best for this experiment to be further carried out on a larger scale. This could be in the form of multiple experiments each looking specifically at one of the three categories: er in the prefield with a single function, er_X in the prefield with an implicit function, or er_X in the prefield with an explicit er occurring in the midfield. Here, more sentences for each condition could be included, increasing the robustness of the results.

Further research is required to investigate whether the quantitative function is required in sentences such as (14) and (15). This could mean looking at sentences that seemingly require the quantitative function by including a headless quantity noun phrase. By comparing such sentences with and without er_Q present in the midfield, a better understanding may be reached.

Furthermore, investigating whether the grammatical function of the quantity in sentences with existential er_{XQ} with an implicit quantitative function in the prefield affects acceptability. The results in this paper suggest a difference in acceptability between the quantity being the subject of the sentence or the object. However, further research is required to determine whether this is accurate.

This paper did not take into account er_X in the prefield with more than one implicit function. This would be an important topic to research in the future. Another point for future research is looking at

the acceptability of the functions of *er* when *er* is not present in the prefield. According to linguistic theories, adjacent occurrences are not possible, but multiple implicit functions are allowed (Neeleman and Van de Koot, 2006; Webelhuth and Bonami, 2019). Testing the acceptability of such sentences for native Dutch speakers would be interesting.

Additionally, further research can be done to determine how strong of an influence educational level or regional dialects have on the acceptability of the functions of *er* in certain positions in the sentence.

References

James Dean Brown. What issues affect likert-scale questionnaire formats. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 4(1), 2000.

J. Bruin. newtest: command to compute new test @ONLINE, February 2011. URL `https://stats.idre.ucla.edu/stata/ado/analysis/`.

James T Croasmun and Lee Ostrom. Using likert-type scales in the social sciences. *Journal of Adult Education*, 40(1):19–22, 2011.

Berna de Boer, Margaret van der Kamp, and Birgit Lijmbach. *Nederlands in gang.* 2010.

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

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.

Jana Häussler and Tom Juzek. Hot topics surrounding acceptability judgement tasks. 2017.

Ad Neeleman and Hans Van de Koot. Syntactic haplology. *The Blackwell companion to syntax*, pages 685–710, 2006.

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

Jon Sprouse and Diogo Almeida. The role of experimental syntax in an integrated cognitive science of language. *The Cambridge handbook of biolinguistics*, 181:202, 2013.

Gert Webelhuth and Olivier Bonami. Syntactic haplology and the Dutch proform "er". In Stefan Müller and Petya Osenova, editors, *Proceedings of the 26th International Conference on Head-Driven Phrase Structure Grammar*, pages 100–119. CSLI Publications, 2019.

A   Appendix

Stimuli sentences used are provided below. Some sentences were taken from Webelhuth and Bonami (2019), Donaldson (2008) and Neeleman and Van de Koot (2006).

er in the prefield with a single function:

- er_X in the prefield
    - Er loopt een man op straat.
    - Er staat een foto van haar in dit boek.
    - Er staat een paard in de gang.
    - Er is veel gebeurd hier.
- er_L in the prefield
    - Er ben ik nooit geweest.
    - Er loop ik naartoe.
    - Er woont ze al tien jaar.
    - Er hebben wij gedanst.
- er_P in the prefield
    - Er speelde haar kat vandaag op de boom.
    - Er liep Harry gisteren overheen.
    - Er kijk ik vaak naar.
    - Er heb ik het brood mee gesneden.
- er_Q in the prefield
    - Er vind ik niks van.
    - Er heeft hij vijf.
    - Er heeft Sara drie gemaakt.
    - Er aten de kinderen vijf in een half uur.

er_X in the prefield with an implicit function:

- er_{XL} in the prefield
    - Ken je de Korreweg? Er gebeuren ongelukken.
    - Ken je Upsilon? Er wonen veel studenten.
    - Ken je de treinstation? Er wacht er een vrouw al huren.

- Ken je Spijkenisse? Er is een brug gebouwd.
- er_{XP} in the prefield
    - Er wordt morgen over gesproken.
    - Er keken veel mensen naar.
    - Er heeft een man brood mee gesneden.
    - Er zat gisteren een kat in.
- er_{XQ} in the prefield
    - Er zijn gisteren twee gestolen.
    - Er lagen drie op tafel.
    - Er bestaan twee.
    - Er heeft iemand vier gebroken.

er_X in the prefield with an explicit er in the midfield:

- er_X in the prefield er_Q in the midfield
    - Er zijn er twee gestolen.
    - Er lagen er gisteren drie op tafel.
    - Er lopen er drie over straat.
    - Er waren er tien verkocht vandaag.
- er_X in the prefield er_P in the midfield
    - Er wordt er morgen over gesproken.
    - Er keken er veel kinderen naar.
    - Er denken er mensen aan.
    - Er helpt er een man mee.
- er_X in the prefield er_L in the midfield
    - Ken je Middelstum? Er wonen er veel boeren.
    - Ken je de Korreweg? Er gebeuren er ongelukken.
    - Ken je de treinstation? Er wacht er een man al dagen.
    - Ken je Spijkenisse? Er is er een brug gebouwd.

Filler sentences

- Correct sentences from de Boer et al. (2010)

  - Uit welk land kon je? Ik kom uit Engeland.
  - We gaan de tekst ook lezen.
  - Hij is jonger, maar wel langer.
  - Nu wil ik ook een thee.
  - We zijn op vrijdag weer thuis.
  - Ik heb problemen met de buren.
  - Kunt u mij helpen? Ik zoek een spijkerbroek.
  - Ik verveel me nooit.
  - Ik houd meer van sport.
  - Mijn broer is jarig.
  - Het is zeker een uur lopen vanaf hier.
  - Ik neem de plattegrond mee.
  - Ik heb iets voor jullie meegenomen.
  - Ik ben wel een beetje vroeg.
  - Ik ben al twee weken verkouden.
  - Hier is het lekker warm.
  - Heb je je sportkleren bij je? Je abonnement gaat vandaag in.
  - Ik kijk even op de computer.
  - Dit is een leuk eetcafe.
  - Zullen we zo naar bed gaan? Ik ben moe.

- Incorrect sentences were translated from Hindi to Dutch

  - Ik twintig het jaar van ben.
  - De jongen het huis op is de.
  - De kinderen de school in de lees zijn.
  - Dat de man daar loop is geweest.
  - Hier drie de mensen zijn.
  - Dat daar vijf het jaar van blijf is geweest.
  - Kat waar is? De kat de boom op beklimmen is geweest.
  - Morgen twee de boeken de diefstal wees was weg.
  - Harry daar velen de mensen aan zie zou kunnen is.
  - Harry ken jij? Hij vijf de vrucht eet zijn.
  - Ik het eten kopen ga ben.
  - Een het meisje speelgoed met spelen is de.
  - De koeien de boerderij in zijn.
  - Heb jij over Grotemarkt gehoord? Ik daar ooit nee weg.
  - Een man hierover zal praten.
  - Hij drie de beker de thee gemaakt.
  - De vrouw aan een het goede de boom kreeg.
  - Twee de honden het deel weg.
  - Sara ken jij? Ik haar een de brief had geschreven.
  - Ik het huis moeten gaan.