university of groningen / faculty of science and engineering

# Regional variance in the use of quantitative "er"

Bachelor's Project Thesis

Nadege Dekker, s3348482, n.dekker.3@student.rug.nl,
Supervisor: Dr. S. Jones

**Abstract:** The Dutch word "er" loosely translates to "there", and is considered an unstressed form of the word "daar" ("there"). The word "er" may therefore be expected to be strictly locative ($er_L$), though in reality it has four possible functions in Dutch. These different uses of "er" are subject to grammatical rules, and while there is some room for interpretation by the speaker, the Dutch are quite specific in how they prefer to use the word "er". This article specifically investigates the use of $er_Q$ by speakers from different regions of the country, if this use agrees with literature, and whether the animacy of the object that "er" refers to makes speakers more or less likely to use "er" in a sentence. It is expected according to literature that the southern provinces will use $er_Q1$ more often than strictly required by grammar rules. 177 participants were asked to fill out a survey. The questions were all multiple choice, giving participants the choice between neither, one or two of the responses to a posed question. Their preferred responses were recorded and analysed by use of an ANOVA (Analysis Of Variance) test. Participants from the northern three provinces used $er_Q$ less frequently than other participants [$F(2, 169) = 20.62, p = 9.69{*}e^{-9}$], which stands in contradiction with literature (Haeseryn, Romijn, Geerts, de Rooij, and van den Toorn (1997)). Furthermore, participants were significantly more likely to use "er" when referring to an inanimate object than an animate object, regardless of which province they inhabited [t(349)=3.53, p=$4.72 * e^{-4}$].

## 1 Introduction

While Dutch people don't give the small word "er" a second thought, they immediately notice when it is used incorrectly. The grammar rules regarding the use of "er" are relatively complex, and the use of "er" that is considered acceptable varies per region according to literature (Haeseryn et al., 1997). To effectively process natural language, and to ensure that a system is robust enough to process speech from different regions, these regional differences first need to be investigated. According to Donaldson (2008) "er" is in origin an unstressed form of "daar", and thus literally means "there". "Er" has four different functions: Locative "er" $er_L$, existential "er" $er_X$, prepositional "er" $er_P$ and quantitative "er" $er_Q$, which will be discussed in their own subsections. It is also possible to have a combination of multiple functions of "er". Only in the combination of the existential "er" and quantitative "er" will there be two "er"s in a sentence.

### 1.1 Four forms of "er"

#### 1.1.1 Locative "er"

Locative "er" translates to "there", and replaces "daar" or "hier" ("there" or "here") in unstressed position (Webelhuth and Bonami, 2019). $Er_L$ cannot occur at the beginning of a sentence because of its weak meaning. Sentence 1 is an example of a sentence containing an $er_L$:

(1) *Het boek heeft er jaren gestaan.*
The book has been there for years.

#### 1.1.2 Existential "er"

Existential "er" can often be translated as "there", though this is not always possible as "er" has no real meaning. Often this "er" can be omitted but is preferred by the Dutch ear. $Er_X$ is the only function of "er" that can happen at the beginning of a sentence (Webelhuth and Bonami, 2019). There are two forms of $er_X$. The first occurs when the subject is indefinite (sentence 2), and the second as

the subject of an impersonal passive sentence (sentence 3) (Donaldson, 2008). Sentence 3 also shows an example of an $er_X$ that is not translated.

(2) *Er lag een boek.*
There was a book.

(3) *Er werd een boek neergelegd.*
A book was put down.

#### 1.1.3 Prepositional "er"

Prepositional "er" occurs in combination with a preposition. Here "er" replaces the pronoun "het" ("it"), as Dutch sentences containing "het" and a preposition are impossible (Voortman, 2005). $Er_P$ can thus often be translated by "it". "Er" and the preposition can be written as one word (sentence 4), but this does not always happen (sentence 5).

(4) *Ik leg het boek erop.*
I'm putting the book on it.

(5) *Het boek ligt er al lang op.*
The book has been on it for a long time.

#### 1.1.4 Quantitative "er"

Quantitative "er" is comparable to the French "en". When translating to English $er_Q$ literally means "of it/them" but is often not literally translated. An $er_Q$ must appear in a certain position, following the finite verb. The use of this form of "er" is a bit more complicated than the other forms, causing it to be more of a struggle to learn, not only for foreigners but also Dutch children (van Hout, Veenstra, and Berends, 2011).
A sentence containing $er_Q$ does not occur as a standalone sentence, as "er" refers back to a preceding discourse, making it an anaphor. The main grammar rule for $er_Q$ is that a sentence containing a numeral or a weak quantifier but omitting a noun phrase needs an er:

(6) *Ik heb er twee.*
I have two (of it).

However when the numeral or weak quantifier is combined with an adjective, the use of $er_Q$ is no longer grammatically correct. Though according to the Algemene Nederlandse Spraakkunst (ANS, Haeseryn et al., 1997) the use of er in a sentence with an adjective is still allowed in practice in the provinces south of the Dutch rivers:

(7) *Ik heb twee gele.*
I have two yellow ones.

It is interesting to note that the Dutch sentence loses the word "er" when an adjective is added, while the English sentence gains an extra word "one(s)" (see sentence 7).
Not all quantifiers warrant the use of "er". E.g. strong quantifiers do not need an "er". A strong quantifier presupposes that other referents exist (De Hoop, 1992 and De Jong, 1983. As cited in Berends, 2019). For example, in the case of the strong quantifier "sommige" (some) in sentence 8, the assumption is that there are "andere" (others). The weak quantifiers do not have this property (sentence 9).

(8) *Ik heb sommige.*
I have some.

(9) *Ik heb er enkele.*
I have a few (of them).

### 1.2 Referential choices

There are many ways one can refer to an item, from long descriptive noun phrases to a short one-word pronoun, like "er". There are many factors that influence one's reference choice:
The longer the time is between mentioning the item and the moment that it is referred to the more likely it becomes that a noun phrase will be used. The kind of reference depends on how mentally available the item is. An item that is more mentally available is more likely to be referred to with a less descriptive reference. The mental availability depends on how likely an item is to occur in the context. An item that has occurred a sentence earlier is more likely to show up again and is therefore more easily accessible. The same goes for an item that is more likely to exist in the context. So a bird in a tree is more accessible than a fish in a tree. The physical accessibility plays no role in the choice of reference. (Ariel, 1990; Chafe, 1994; Givón, 1983. As cited in Vogels, 2014)
Animacy of the referred item also has an effect on whether a pronoun or noun phrase is used. Fukumura and van Gompel (2010) found that people were more likely to refer to an item with a pronoun when the item was animate than when it was inanimate. A limitation of this study was that they

only looked at sentences where the reference was the subject of a sentence.
Vogels, Krahmer, and Maes (2013) looked to confirm the study of Fukumura and van Gompel (2010) with Dutch participants. They also looked at the influence of animacy on the reference choice when the reference took a non-subject function. They then found that Dutch speakers were also more likely to refer to the item with a pronoun when the item was animate rather than inanimate.
However they also found that when the reference did not take the subject position in a sentence, participants were more likely to use a pronoun to refer to the item when it was inanimate rather than animate. It is important to note that animate and inanimate are not two completely separate categories, the transition is more like a gradient where things are perceived from very animate to very inanimate and in between. For example humans are perceived as more animate than animals, animals as more animate than cars and cars as more animate than books.

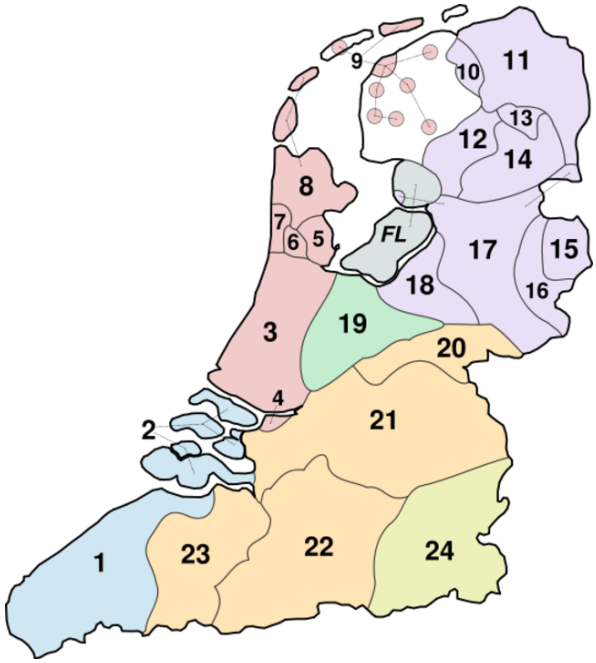


**Figure 1.1: Dutch dialects grouped by Jo Daan in 1968. (Image taken from Wikimedia Commons)**

## 1.3 Dialects

Dialects also influence how and when "er" is used. It is hard to find a language that has no dialects. A dialect is a form of the language which is specific for a group of people. This grouping can be related to a geographical region or social class. The most noticeable difference between dialects is their different words for a similar item. But dialects also might differ in grammar and semantic meaning (Stroop, 2012). The different dialects' features have their own boundaries. These boundaries are called isoglosses. Dialects are a result of speakers interacting and adopting language features from one another other. As a result dialects on separate sides of a border might be more similar than two dialects within the same country, as they have a smaller spatial separation. Figure (1.1) shows a map of the Dutch language region, where different colors denote different dialect groups, and the numbers denote different dialects. However, even within the dialects there might be smaller sub dialects.

## 1.4 Research Aim

Two potential relations will be investigated in this paper: Firstly, the influence of region on the use of $er_Q$, and secondly, the influence of animacy of the referred object on the use of "er". As previously discussed, based on the grammar rules described in the ANS, it is expected that in sentences containing a numeral but omitting a noun phrase an "er" will be needed, and that sentences containing a numeral and omitting a noun phrase but containing an adjective will not need an "er", with the exception of the South up to the rivers. This leads to the first research question: *Does the use of quantitative "er" vary based on region?* With the hypothesis being that there will be no significant difference in the use of "er" in basic sentences, but that "er" will be used more often in the South for the adjective sentence.
The second research question is: *Does the animacy of the referred object influence the use of "er"?* This will be considered in sentences both with and without an adjective as these have different prediction for the use of "er". Based on the findings of Vogels et al. (2013) the expectation is that "er" will be used less in sentences where "er" refers to an animate object, as $er_Q$ can not be the subject of the sentence. To find an answer to the two research

questions a survey will be carried out among native Dutch participants from everywhere in the Netherlands.

# 2 Methods

## 2.1 Participants

In total, 177 participants finished the experiment. The participants all grew up in the Netherlands. The participants were split into four age groups: There were five participants younger than 18, 90 participants were between 18 and 31, 48 participants were between 31 and 50 and 39 participants were older than 50. No data on gender or bilingualism was collected. The participants were asked to volunteer in the experiment via a message that was spread on social media. No compensation was offered.

## 2.2 Survey

The experiment was created as an online survey using Qualtrics. The entire survey was in Dutch. The survey began with a greeting and instructions of how to answer the questions in the survey, together with an example question. After the example the participants were asked to fill in their age group and the province they grew up in. Then they were asked 50 questions in a random order. Of these questions 20 were target questions and 30 were filler questions. The questions consisted of a question sentence with two answer options; A and B. The participants were then asked to choose which sentence they used in their day-to-day language. They could answer: "sentence A", "sentence B", "both sentences" and "neither sentence".
The choice for a multiple choice between sentences instead of acceptability judgement task was made based on a discussion in Sprouse and Almeida (2013). Where they discussed how a forced choice task did show significant results while an acceptability task did not. Multiple choice was chosen as a compromise between the two, to force participants to consider both sentences, while giving them the choice to express dislike or preference for both sentence. This option was added to investigate if there is a transition region, where both sentence are considered acceptable.

### 2.2.1 Target Q&A pairs

To examine the effect of region on the use of "er", four types of Q&A pairs were constructed. They follow the same structure, starting with a question followed by answer options A and B. The sentence structure of the answer options is as short as possible, containing a subject, a verb, ($er_Q$) and a numeral or a weak adverb of quantity for the simple sentences. Where option A is a sentence containing an $er_Q$ and option B a sentence without $er_Q$.
Two things were manipulated about these sentences; the answer options could either have an adjective at the end of the answer options or not. And $er_Q$ could either refer to an animate or an inanimate object. This leaves us with the following types of Q&A pairs. The animate basic Q&A pair (10), the inanimate Q&A pair (11), the animate adjective Q&A pairs (12) and lastly the inanimate adjective Q&A pairs (13). Each example (10 through 12) Q&A pair is one of five Q&A pairs in a group.

(10) *Hoeveel vogels zag jij?*
   a. *Ik zag er vijf.*
   b. *Ik zag vijf.*

   How many birds did you see?
   a. I saw *er* five.
   b. I saw five.

.

(11) *Hoeveel appels at jij?*
   a. *Ik at er drie.*
   b. *Ik at drie.*

   How many apples did you eat?
   a. I ate *er* three.
   b. I ate three.

(12) *Hoeveel vogels zag jij?*
   a. *Ik zag er vijf zwarte.*
   b. *Ik zag vijf zwarte.*

   How many birds did you see?
   a. I saw *er* five black ones.

b. I saw five black ones.

(13) *Wat voor appel at jij?*
a. *Ik at er drie groene.*
b. *Ik at drie groene.*

What kind of apple did you eat?
a. I ate *er* three green ones.
b. I ate three green ones.

When constructing the Q&A pairs special care was taken to avoid the use of strong adverbs of quantity and to make sure $er_Q$ was unambiguous.

#### 2.2.2 Filler sentences

The filler sentences consist of two kinds of filler sentences, with 20 grammatically correct sentences and 10 grammatically incorrect sentences. First there are the grammatically correct filler sentences, these consist of sentences where the answer options contain other variants in the Dutch language. An example is:

(14) *Hoe laat is het?*
a. *Het is twintig over drie.*
b. *Het is tien voor half vier.*

What is the time?
a. It is twenty past three.
b. It is ten before half past three.

Then there are the 10 grammatically incorrect sentences. These are included to detect uncooperative participants and thus filter out noise in the collected answers. The grammatically wrong sentences have answer options that contain obvious grammatical errors. There are also five grammatically wrong sentences that contain an "er", so the participants would not assume that all sentences containing "er" must be correct. An example is:

(15) *Hoe groot is Sandra?*
a. *Ze is groter als ik.*
b. *Ze is groter dan mij.*

How tall is Sandra?
a. She is taller as I.
b. She is taller than me.

All sentences used in the experiment can be found in Appendix A.

### 2.3 Grouping of the data

To be able to analyze the data, the answers of the participants have to be ordered into two categories: true or false to prediction. As the basic Q&A pairs have a different prediction for the use of "er" than the adjective Q&A pairs (see introduction), the data is grouped differently for the two types of Q&A pairs. The prediction for the basic Q&A pairs is that the participant will prefer to use *er*. Therefore, the two answer options containing this preference for "er" are grouped together under true, and the other two under false for the basis Q&A pairs. The same goes for the adjective Q&A pairs, where the prediction is that the answer option without "er" will be preferred. The answer options containing a preference for the answer without "er" are grouped under true, and the other two options under false. You can see the grouping in table 2.1.

The provinces, of which a map can be found in the appendix C, are grouped together into three regions: North, Middle and South. This is done to make the size of the different groups bigger and thus gain statistical power. The southern group consists of the provinces Limburg, North-Brabant and Zeeland. These are the provinces in which a difference is expected to be observed, as the ANS (Haeseryn et al., 1997) tells us that this difference in language exists in the southern regions up to the rivers. The North is grouped as Friesland, Groningen and Drenthe. The last group will be referred to as the Middle group. This consists out of North- and South-Holland, Overijssel, Gelderland and Utrecht. Though figure (1.1) shows that the province of Overijssel is considered to be in the Nedersaksen region, the dialect groups (17 & 18) in Overijssel directly border the Brabants dialect region (20 through 23). Therefore, a gradient transition region that affects dialect groups 16 through 18 is expected to form between the two groups, thus it is expected that the province of Overijssel will agree more with the middle group than the north group.
The province Flevoland is not used in the analysis as this province is relatively new, its creation was finished in 1968. Therefore, its inhabitants come from other provinces, thus Flevoland doesn't have

its own dialect.

# 3 Results

None of the data were excluded from the results based on the participants answers on the control Q&A pairs. This judgement is based on the fact that almost all participants answered at least 50 per cent of the control Q&A pairs with both answers being incorrect. The participants who answered more control Q&A pairs as correct often also commented that they believed the control Q&A pairs contained accidental mistakes, and they thus chose to answer them as normal Q&A pairs. This did not seem to affect the answers they gave on the target Q&A pairs.

## 3.1 Basic Q&A pairs

The results for the basic Q&A pairs are shown in figures (3.1, 3.2). Figure (3.1) shows the mean number Q&A pairs answered with a preference for the use of "er". This graph illustrates that the participants from the three northern provinces (left) appear to have less of a preference for the use of "er" than participants from the other two regions.

To investigate the correlation between region and preference for the use of "er", the data were grouped by region, in accordance with section 2.3, and placed in a box-and-whiskers plot. Figure 3.2 shows the mean, variance, error and outliers of the basic Q&A pairs.

A one-way between subjects ANOVA (analysis of variance) was carried out to see if there was an effect of region on the preference for "er". The ANOVA test found that there was a significant difference between the regions at the $p < .05$ level $[F(2, 169) = 20.62, p = 9.69 * e^{-9}]$.
Post hoc comparisons using the Tukey HSD test then indicated that the North differed significantly from the South (p=0.001) and the Middle (p=0.000).

## 3.2 Q&A pairs with an adjective

The questionnaire also contained Q&A pairs containing an adjective. Figure (3.3) implies that there is not as strong a correlation between the use of "er" and region when an adjective is used in the sentences.

A one-way between subjects ANOVA (analysis of variance) was carried out to make sure there was indeed no effect of region on the preference against the use of "er". The ANOVA test found that there was indeed no significant difference between the regions at the $p < .05$ level $[F(2, 169) = 0.466, p = 0.628]$.

## 3.3 Animacy

The questionnaire contained Q&A pairs that were either animate or inanimate by design. Figure (3.5) shows the mean use of "er" per province, and differentiates between animate and inanimate sentences respectively. The participants from all but one province (Zeeland) used "er" less in animate sentences than inanimate sentences. To investigate this apparent preference, the answers from all participants were grouped not by region, but rather by animacy of the question. Figure (3.6) shows a box-and-whiskers plot of the answers according to animacy.
Analysis of these data gives us a significant difference in the preference for "er" between animate (M=4.61, SD=1.87) and inanimate (M=5.23, SD=1.71) Q&A pairs; t(349)=3.53, p=$4.72 * e^{-4}$. There was found to be a difference between the animate (M=4.12, SD=1.23) and inanimate (M=3.79, SD=1.02) sentences in the adjective Q&A pairs (t(340)=2.82, p=0.01). No significant difference was found between animate(M=3.98, SD=1.30) and inanimate (M=4.13, SD=1.19) for the basic Q&A pairs on their own; (t(348)=1.10, p=0.26).

# 4 Discussion

Two specific Q&A pairs stood out from the other Q&A pairs, as participants answered these two in disagreement with the grammar rules relatively more often. The two questions, and their respective unexpected answers that were given, are as follows:

(16) Q: *Hoeveel deelnemers waren er aanwezig?*
A: *Er waren zeventien.*

Adjective Q&A pair. Prediction "er" will not be used (ER-).

Basic Q&A pair. Prediction "er" will be used (ER+).

| | True | False |
|---|---|---|
| True | ER+ and ER- true (both sentence are correct) | ER+ false and ER- true (sentence B) |
| False | ER+ true and ER- false (sentence A) | ER+ and ER- false (neither sentence is correct) |

**Table 2.1: The grouping of different answer options for the two Q&A pair types**

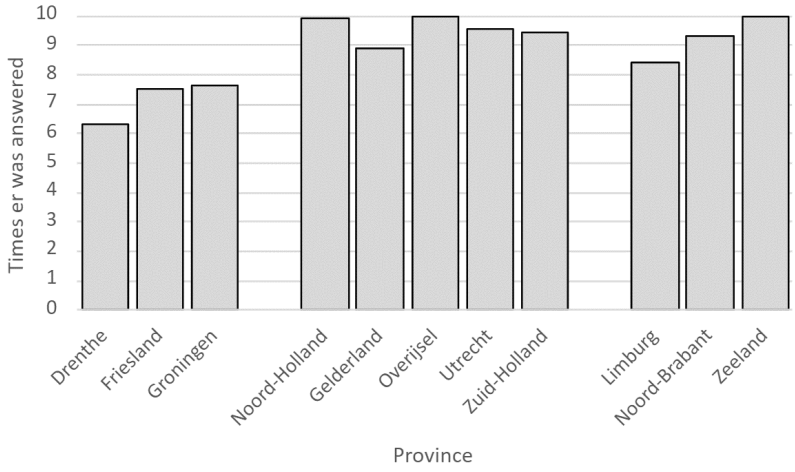


**Figure 3.1: Bar graph showing the mean of basic Q&A pairs answered according to the prediction (use of "er" is preferred), out of the ten basic Q&A pairs. Showing the individual provinces and their grouping by region.**

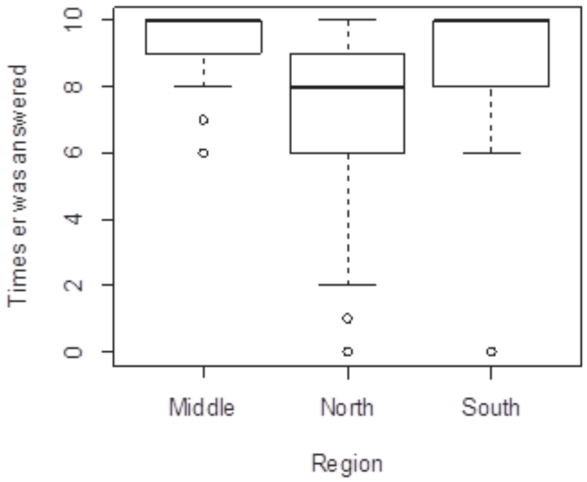


**Figure 3.2: Box plot showing the mean of the of basic Q&A pairs answered according to prediction (use of "er" is preferred). It also shows the standard deviation (box), the error (bars) and the outliers (dots).**

Q: How many participants were present?
A: There were seventeen.

(17) Q: *Heeft u nog knopen?*
A: *Ik heb er een heleboel kleine.*

Q: Do you still have buttons?
A: I have a bunch of small ones.

The answer shown for example 16 is grammatically incorrect, as the question does not contain an adjective or strong quantifier and does contain a numeral. There is already an "er" present in the sentence, $er_X$, however "er" can not have the double function $er_X$ and $er_Q$ as discussed in the introduction. Thus the answer requires the use of "er" twice, one $er_X$ and one $er_Q$. However, many participants responded according to the example. This may be because they gave "er" a double function (existential and quantitative). However, in another question (18) containing both $er_X$ and $er_Q$ in the answer options, participants did answer according to the grammar rules.

(18) Q: *Zijn er nog andere appels?*
A: *Ja, er zijn er nog twee.*

Q: Are there still other apples?
A: Yes, there are still two.

Similarly, relatively many participants gave the grammatically incorrect answer to example (17), though I could not discern why. These two Q&A pairs may have skewed the results of the animacy statistics as examples (16) and (17) happen to be animate and inanimate respectively, and were answered differently than most other questions. Future research should consider how to avoid such ambiguous results by identifying which Q&A pairs are outliers and accounting for these outliers in statistical analysis. It might also be worth investigating why these Q&A pairs were not answered according to the grammar rules.

It was expected that participants from the southern regions of the Netherlands would use "er" in adjective sentences more often than the rest of the Netherlands. This was not reflected by the questionnaire responses. The south was expected to use "er" more often, as in Flemish the use of "er" in adjective sentence is grammatically correct. This manner of speaking was expected to diffuse into the southern regions of the Netherlands. As this manner of speaking was not observed in the southern regions, it might be interesting to expand the investigation of the use of "er" among native Dutch speakers to include the Flemish.

Multiple participants commented anonymously that the questionnaire contained grammatically incorrect questions and answers, thereby indicating they had not read or misunderstood the instructions which stated that they should not select answers that do not reflect their everyday speech. All sentences that participants specifically referred to as "grammatically wrong" were control sentences. These were intended to ensure that participants read and understood all Q&A pairs completely, and had now lost their function. Future research may find it worthwhile to inform participants that the questionnaire contains control sentences, and to instruct them how to respond explicitly, to avoid participants overthinking their answers. Even if the results of the questionnaire may therefore be slightly less resistant to participants selecting random responses, the validation of responses would be greatly facilitated. Nonetheless, most participants answered more than 50 per cent of the control sentences correctly, and so no results were discarded

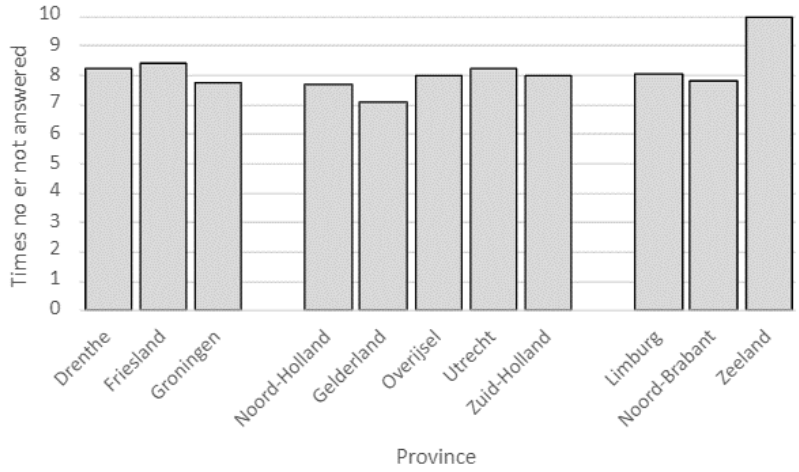


**Figure 3.3: Bar graph showing the mean of adjective Q&A pairs answered according to the prediction (use of no "er" is preferred), out of the ten adjective Q&A pairs. Showing the individual provinces and their grouping by region.**

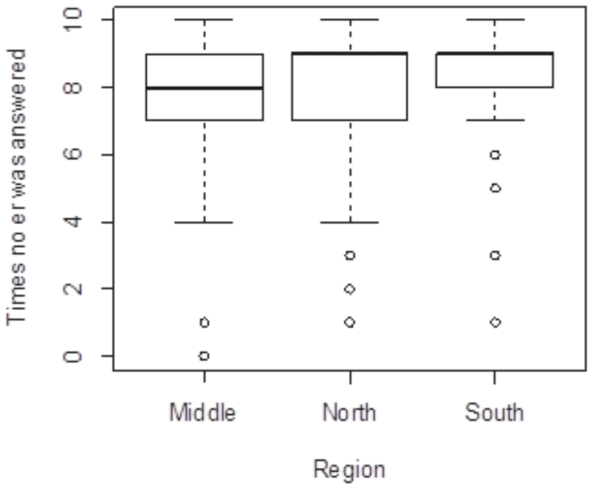


**Figure 3.4: Box plot showing the mean of the of adjective Q&A pairs answered according to prediction (use of no "er" is preferred). It also shows the standard deviation (box), the error (bars) and the outliers (dots).**

by answers given in response to the control questions.

The age of participants was determined not to be a factor in the use of "er". Figure (4.1) shows a box plot of the total amount of times participants preferred the sentence containing "er" averaged per age group. Table 4 shows the average number of times each answer type was answered. Averaged over all Q&A pairs the answers A and B are approximately equally likely to be given, and participants were also roughly equally likely two answer "both" or "neither" regardless of age. The youngest age group may seem to have had a preference for option A, but this is not statistically significant due to the group only containing five participants.

There was a relatively large in-group variation,

| **Age group** | **A** | **B** | **A and B** | **Neither** |
|---|---|---|---|---|
| **<18** | 10.80 | 8.60 | 0.20 | 0.40 |
| **18-30** | 7.32 | 7.28 | 1.32 | 4.08 |
| **31-50** | 7.88 | 8.46 | 0.63 | 3.04 |
| **51+** | 8.28 | 9.13 | 0.50 | 2.10 |

both for the regional variance and the influence of animacy questions. But this was expected as the grouping of the provinces resulted in many different local dialects being grouped together. Regarding the influence of animacy, the preference for "er" in inanimate sentence is visible in all 11 provinces.

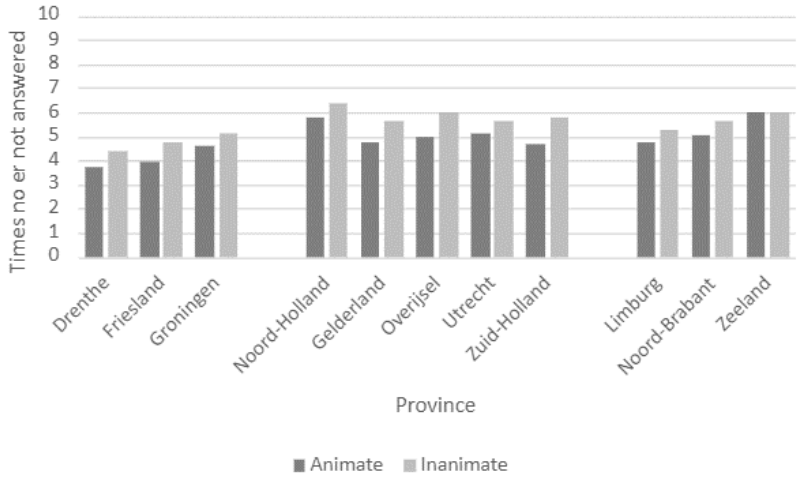


**Figure 3.5: Bar graph showing the mean of Q&A pairs answered with a preference for "er", out of the ten animate and the ten inanimate Q&A pairs. Showing the individual provinces and their grouping by region.**

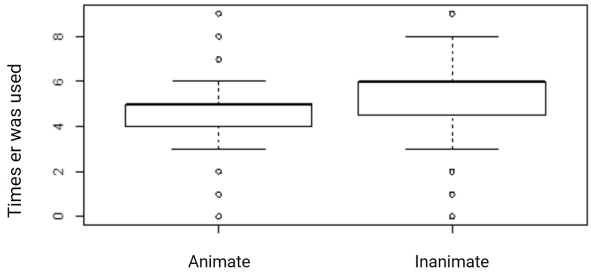


**Figure 3.6: Box plot showing the mean of the of the animate and inanimate Q&A pairs answered with a preference for "er". It also shows the standard deviation (box), the error (bars) and the outliers (dots).**

The probability of this apparent relation being a false positive for all 11 provinces is so small that the result is considered significant.

## 4.1 Part of dialect

The observed variance in the use of "er" does not appear to be governed by a dialect family or by bordering foreign languages. If the effect were to be apparent for an entire dialect family, the whole Nedersaksen region would be expected to agree on the use of "er". Overijssel, despite falling in the region of the Nedersaksen dialects, does not agree

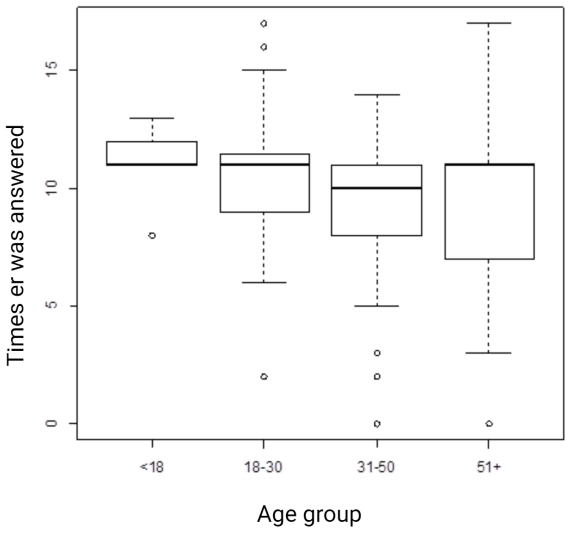


**Figure 4.1: The amount of times participants preferred the sentences containing "er", divided by age group. It also shows the standard deviation (box), the error (bars) and the outliers (dots).**

with the other provinces that are found in this region, namely Groningen and Drenthe. Friesland on the other hand does agree with the two latter provinces, though Friesland has its own language and dialects. Therefore it is possible that this use of "er" in Groningen and Drenthe originates in Fries, rather than Nedersaksen. It may be worth investigating where the isoglosses for the use of $er_Q$ are located. It is interesting to note that the northern region in question also agrees on how to describe the times "twenty past the hour" and "twenty until the hour", where Groningen and Drenthe agree with Friesland that these times should be described as "twintig over ..." (twenty past ...) and "twintig voor ..." (twenty before ...) respectively. The other provinces (Limburg excluded) prefer the use of the expressions "tien voor half ..." (ten before half (before) ..) and "tien over half ..." (ten past half (before) ...), see Appendix B.

# 5 Conclusions

To improve the foundation upon which natural language processing is built, it was investigated whether there is regional variance in the use of $er_Q$, whether this corresponds to the grammar rules described in the ANS, and whether the animacy of the referred object has an influence on the use of $er_Q$. To this end a survey was distributed under native Dutch speakers. 177 participants were asked to fill out a 50 question survey. 20 of these questions pertained to the use of "er". The questions were all multiple choice, giving participants the choice between neither, one or two of the responses to a posed question.

Literature suggested that $er_Q$ would be used more often in the southern provinces than grammatically allowed. According to the results described in this article the ANS (Haeseryn et al. (1997)) incorrectly describes the regional use of "er". Rather than the participants from the south being more likely to use "er" in sentences where grammar rules prescribed they should not, the participants from the northern provinces were less likely to use $er_Q$ in sentences where grammar rules prescribed they should.

Literature predicted that participants would more readily refer to an inanimate item with a pronoun when this reference is not the subject of the sentence. Therefore it was expected that participants would prefer to refer to inanimate items with "er", as $er_Q$ can not be the subject of a sentence. This assumption was confirmed, with the results showing a greater difference in the use of "er" between animate and inanimate Q&A pairs when the answer options contained an adjective. In conclusion this research disagrees with literature on the regional use of $er_Q$, and agrees with the prediction based on literature (Vogels et al. (2013)) on the preferred use of "er" in inanimate sentences. This knowledge may be used in natural language processing to improve the processing of the word "er".

# References


H. Bennis. Hoeveel dialecten heeft het nederlands? `http://www.taalcanon.nl/vragen/hoeveel-dialecten-heeft-het-nederlands/`. [Online; last accessed 05/07/2021].

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

S. Berends. Acquiring dutch quantitative er. Master's thesis, University of Amsterdam, 2019.

A. Blom. Het kwantitatieve er. *Spektator 6*, pages 387–395, 1975.

B. C. Donaldson. *Dutch: a comprehensive grammar*. Abingdon: Routledge, 2008.

K. Fukumura and R. P. G. van Gompel. The effect of animacy on the choice of the referring expression. 2010.

S. Grondelaers, D. Speelman, and A. Corbonez. Regionale variatie in de postverbale distributie vanpresentatief er. 2001.

S. Grondelaers, D. Speelman, and D. Geeraerts. National variation in the use of er "there".regional and diachronic constraints on cognitive explanations. *Journal of Experimental Psychology-learning Memory and Cognition*, pages 153–205, 2008.

W. Haeseryn, K. Romijn, G. Geerts, J. de Rooij, and M.C. van den Toorn. *Algemene Nederlandse Spraakkunst, Tweede, geheel herziene druk*. Groningen/Deurne, Martinus Nijhoff uitgevers/Wolters Plantyn, 1997.

S. M. Jones. In M. Butt and I. Toivonen, editors, *Proceedings of the LFG'20 Conference, On-Line*, page 148–168. Stanford,CA: CSLI Publications., 2020.

A. Neeleman and H. van de Koot. Syntactic haplology. In M. Everaert and H. van Riemsdijk, editors, *The Blackwell Companion to Syntaxe*, volume 1, pages 685–710. London, Blackwell, 2006.

J. Sprouse and D. Almeida. The role of experimental syntax in an integrated cognitive science of language. In C. Boeckx and K. K. Grohmann, editors, *The Cambridge Handbook of Biolinguistics*, pages 181–202. Camebridge, CUP, 2013.

J. Stroop. De dialecten van het nederlands, 2012. URL `https://www.nemokennislink.nl/publicaties/de-dialecten-van-het-nederlands/`.

A. van Hout, A. Veenstra, and S. Berends. All pronouns are not acquired equally in dutch: Elicitation of object and quantitative pronouns. In M. Pirvulescu, M. C. Cuervo, Perez-Leroux. A. T., J. Steele, and N. Strik, editors, *Selected Proceedings of the 4th Conference on Generative Approaches to Language Acquisition North America (GALANA 2010)*, pages 106–121. Cascadilla Proceedings Project, Somerville, MA, USA, 2011.

J. Vogels. Referential choices in language production. Master's thesis, Tilburg University, 2014.

J. Vogels, E. Krahmer, and A. Maes. When a stone tries to climb up a slope: the interplay between lexical and perceptual animacy in referential choices. *frontier in psychology*, 2013.

W. Voortman. The use of er. `https://www.dutchgrammar.com/_word_docs/Er.pdf`, 2005. [Online; last accessed 05/07/2021].

G. Webelhuth and O. Bonami. Syntactic haplology and the dutchproform "er". In S. Muller and P. Osenova, editors, *Proceedings of the 26th International Conference onHead-Driven Phrase Structure Grammar*, pages 100–119. CSLI Publications, 2019.

Wikimedia Commons. Dutch dialects in the netherlands, flanders and french-flanders. `https://commons.wikimedia.org/wiki/File:Dutch-dialects.svg`. [Online; last accessed 05/07/2021].

# A Appendix

| Question | Answer "A"<br>Including "er" in the "er" related questions | Answer "B"<br>Not including "er" in the "er" related questions |
|---|---|---|
| **Basic sentences** | | |
| Hoeveel vogels zag jij?* | Ik zag er vijf | Ik zag vijf |
| Hoeveel kinderen heeft u?* | Ik heb er drie | Ik heb drie |
| Hoeveel deelnemers waren aanwezig?* | Er waren er zeventien | Er waren zeventien |
| Hoeveel koeien heeft de boer?* | Hij heeft er honderd | Hij heeft honderd |
| Hoeveel pinguïns zijn er in Antarctica?* | In Antarctica zijn er 40 miljoen | In Antarctica zijn 40 miljoen |
| Hoeveel appels at jij? | Ik at er drie | Ik at drie |
| Heeft u al een spaarkaart? | Ik heb er twee | Ik heb twee |
| Heb je een paar knopen voor mij? | Ik heb er een heleboel | Ik heb een heleboel |
| Zijn er nog andere appels? | Ja, er zijn er nog twee | Ja, er zijn nog twee |
| Hoeveel plantensoorten kent u? | Ik ken er tweendertig | Ik ken tweendertig |
| **Sentences with an adjective** | | |
| Hoe veel vogels zag jij?* | Ik zag er vijf zwarte | Ik zag vijf zwarte |
| Hoe veel vissen heeft u?* | Ik heb er vijf verschillende | Ik heb vijf verschillende |
| Wat voor koeien heeft de boer?* | Hij heeft er honderd bruine | Hij heeft honderd bruine |
| Heeft u buitenlandse familieleden?* | Ik heb er drie Amerikaanse | Ik heb drie Amerikaanse |
| Heeft u gevaarlijke slangen?* | Ja, ik heb er twee giftige | Ja, ik heb twee giftige |
| Wat voor appels at jij? | Ik at er drie groene | Ik at drie groene |
| Heeft u al een spaarkaart? | Ik heb er twee volle | Ik heb twee volle |
| Heb je een paar knopen voor mij? | Ik heb er een heleboel kleine | Ik heb een heleboel kleine |
| Zijn er nog andere appels? | Nee, er zijn er geen andere | Nee, er zijn geen andere |
| Wat voor pizza's heeft u? | Ik heb er grote en kleine | Ik heb grote en kleine |

| Question | Answer "A"<br>Including "er" in the "er" related questions | Answer "B"<br>Not including "er" in the "er" related questions |
|---|---|---|
| **Filler sentences** | | |
| Hoe laat is het? | Het is twintig over drie | Het is tien voor half vier |
| Hoe lang moeten we nog rijden? | Nog één uur en drie kwartier | Nog één uur en vijfenveertig minuten |
| Welk broodje wil je? | Ik wil graag een bolletje | Ik wil graag een kadetje |
| Hoe veel sinaasappels wilt u? | Ik wil graag vijfhonderd gram. | Ik wil graag een halve kilo. |
| Waar bent u geboren? | Ik ben geboren in s'Gravenhage | Ik ben geboren in Den Haag |
| Wat zullen we eten? | Ik heb zin in friet | Ik heb zin in patat |
| Wat wil je drinken | Ik wil graag ranja | Ik wil graag limonade |
| Wat heeft u een apart accent, | Waar kom je vandaan? | Waar kom je weg? |
| Ga je mee uit vanavond? | Nee, ik lig al in bed | Nee, ik lig al op bed |
| Kan ik achterom komen? | Ja, de deur is open | Ja, de deur is los |
| Wat zoemt daar zo? | Oh, er zit een mug in de kamer | Oh, er zit een vlieg in de kamer |
| Wow, wat een mooie jas. | Dankjewel, die heb ik van mijn vader gekregen | Dankjewel, die heb ik van mijn vader gehad |
| Nee dank u, | Ik ben geen fan van zulke dingen | Ik ben geen fan van zo'n dingen |
| Kent u de band Racoon? | Ik heb er slechts van gehoord | Ik heb er enkel van gehoord |
| Lang niet gezien, | Hoe gaat het? | Hoe is het? |
| Waar is Joost? | Hij zit achter de computer | Hij zit op de computer |
| Wil je je schoenen uit doen, | Ik heb net stof gezogen | Ik heb net gestofzuigd |
| Waar hoor je te skeeleren? | Dat mag op de stoep | Dat mag op het voetpad |
| Wat heeft de boer op zijn land? | Hij verbouwt mais | Hij teelt mais |
| We moeten echt eens samen koffie drinken. | Schikt woensdag jou? | Komt woensdag jou uit? |
| **Filler sentences grammatically and/or semantically wrong** | | |
| Wat staat er in de wei? | Een paard er staat | Er een paard staat |
| Hoeveel benzine wil je? | Ik wil er vijf | Ik wil vijf |
| Wilt u liever rijst of pasta? | Ja, graag | Nee, dank u |
| Waar gaat u heen? | Ik rijst op Azie aan | Ik rijst naar Azie. |
| Wat wil je later worden? | Ik wordt leraar | Ik wordt docent |
| Waar is Bob? | Bob zit onder de computer | Bob zit in de computer |
| Hoeveel kinderen heeft u? | Ik heb er twee jongens | Ik er heb twee jongens |
| Hoe lang moet ik doorgaan? | Tot er klaar is | Tot er af is |
| Wat voor sport doe je? | Ik zit er op | Ik doe er aan |
| Hoe groot is Sandra? | Ze is groter als ik | Ze is groter dan mij |

# B Appendix

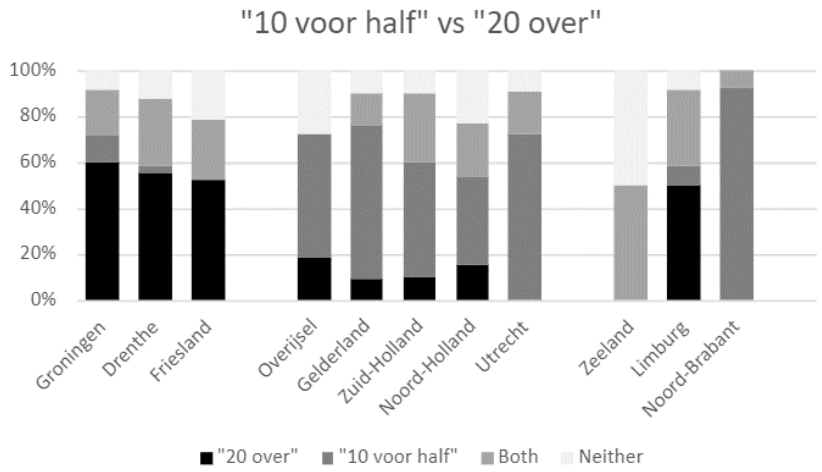


**Figure B.1: Regional use of the phrases: ”twintig over/voor” and ”tien voor/over half” as indicators of time**

# C Appendix


**Figure C.1: Map of Dutch provinces, taken from http://www.pinkgron.nl/nlkaartklein.php**