Title
Split Pronominal Adverbs in Dutch: Searching for Meaning in what is Considered Meaningless

Permalink
https://escholarship.org/uc/item/2qn6h0xf

Author
de Carlo, Nickolas Sebastiano

Publication Date
2017

Peer reviewed|Thesis/dissertation
Split Pronominal Adverbs in Dutch:
Searching for Meaning in what is Considered Meaningless

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Germanic Languages

by

Nickolas Sebastiano de Carlo

2017
ABSTRACT OF THE DISSERTATION

Split Pronominals in Dutch:
Searching for Meaning in what is Considered Meaningless

by

Nickolas Sebastiano de Carlo

Doctor of Philosophy in Germanic Languages

University of California, Los Angeles, 2017

Professor Robert S. Kirsner, Chair

Pure synonymy is a myth. This is true in the lexicon of a language as well as its grammar. As such, why do grammars of Dutch label the following two forms below as synonymous?

Unsplit:   Daarmee heb ik een probleem.
           Therewith have I a problem.

Split:     Daar heb ik een probleem mee.
           There have I a problem with.
           "I have a problem with that."

Calling these two forms above "the same" is an oversimplification. In my dissertation, I test for difference in meaning between the forms and the motivation to choose one form over the other. To accomplish this goal, I use a corpus-based approach integrated with tests from Construction Grammar. We analyze these data in R (a statistics computing language) and Excel. This dissertation includes four tests, all complete. Preliminary findings underline a difference in meaning and provide motivation for the two forms. This is significant, as the current state of
research suggests no real difference or meaning. Only one study, Honselaar (1994), discusses any possible variation in meaning. Additionally, emerging studies in linguistics frequently utilize larger data sets. This dissertation supports the shift toward more quantifiable data coupled with trained qualitative analysis. The quantifiable tests show a meaning difference, questioning the traditional analysis. Unsplit forms exhibit a shift in discourse, and split show continuation thereof.

Regarding motivation for the two forms, the data show clauses with verbs of low use frequency tend to be unsplit, suggesting that higher frequency verbs are more likely to occur with split forms. Frequency tends to be understood and decoded quicker (Fenk-Oczlon, 2001), and thus the listener/reader is generally able to parse the second half of the pronominal adverb, namely the preposition, before its utterance. This suggests that it would be easier to continue forward in the discourse if the form is split. These findings underline the growing need for further development of corpus-based linguistics studies. They welcome the movement of linguistics toward more empirical approaches to language.
This dissertation of Nickolas Sebastiano de Carlo is approved.

Christopher M. Stevens

Hongyin Tao

Robert S. Kirsner, Committee Chair

University of California, Los Angeles

2017
ازیزانم
Table of Contents

1 Problem................................................................................................................................. 1
  1.1 Object of Inquiry .................................................................................................................. 4
  1.2 Function of Pronominal Adverbs ....................................................................................... 6
  1.3 Choice of Pronominal Adverb: Daarmee .............................................................................. 7
  1.4 Met ....................................................................................................................................... 8
  1.5 Daar ..................................................................................................................................... 9
  1.6 Summary of Study ............................................................................................................. 11

2 Construction Grammar ......................................................................................................... 12

3 Source Material (Corpus discussion) ..................................................................................... 15

4 Introduction to the Tests ....................................................................................................... 18
  4.1 Previous Studies into Split Pronominal Adverbs ............................................................... 19

5 Test One: Phrasal Verbs ....................................................................................................... 26
  5.1 Introduction ....................................................................................................................... 26
  5.2 Problem ............................................................................................................................. 28
  5.3 Methodology ...................................................................................................................... 29
  5.4 Collostructional Analysis ................................................................................................. 32
  5.5 Results ............................................................................................................................... 34
  5.6 Individual Verbs ............................................................................................................... 36
  5.7 Closing Discussion ........................................................................................................... 38

6 Test Two: Frequency ............................................................................................................. 42
  6.1 Introduction ....................................................................................................................... 42
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2</td>
<td>Methodology</td>
<td>44</td>
</tr>
<tr>
<td>6.3</td>
<td>Genres/Lists</td>
<td>45</td>
</tr>
<tr>
<td>6.4</td>
<td>Correlation</td>
<td>49</td>
</tr>
<tr>
<td>6.5</td>
<td>Conclusion Test Two</td>
<td>55</td>
</tr>
<tr>
<td>7</td>
<td>Test Three: WordNet &amp; Transitivity</td>
<td>57</td>
</tr>
<tr>
<td>7.1</td>
<td>Methodology</td>
<td>58</td>
</tr>
<tr>
<td>7.2</td>
<td>Data Preparation</td>
<td>60</td>
</tr>
<tr>
<td>7.3</td>
<td>Results</td>
<td>61</td>
</tr>
<tr>
<td>7.4</td>
<td>Transitivity</td>
<td>67</td>
</tr>
<tr>
<td>7.5</td>
<td>Meaning Difference</td>
<td>73</td>
</tr>
<tr>
<td>8</td>
<td>Test Four: Literature &amp; Semantic Integrality</td>
<td>77</td>
</tr>
<tr>
<td>8.1</td>
<td>Methodology</td>
<td>78</td>
</tr>
<tr>
<td>8.2</td>
<td>Implementation of Test Three on the Novels</td>
<td>81</td>
</tr>
<tr>
<td>8.3</td>
<td>Semantic Integrality</td>
<td>83</td>
</tr>
<tr>
<td>8.4</td>
<td>Extracts</td>
<td>86</td>
</tr>
<tr>
<td>8.5</td>
<td>De donkere kamer van Damocles</td>
<td>87</td>
</tr>
<tr>
<td>8.6</td>
<td>Nooit meer slapen</td>
<td>101</td>
</tr>
<tr>
<td>8.7</td>
<td>Review</td>
<td>114</td>
</tr>
<tr>
<td>9</td>
<td>Conclusion</td>
<td>117</td>
</tr>
<tr>
<td>10</td>
<td>Appendix A: Tree Tagger Numerical Values</td>
<td>121</td>
</tr>
<tr>
<td>11</td>
<td>Appendix B: Met Verbs</td>
<td>123</td>
</tr>
<tr>
<td>12</td>
<td>Appendix C: Prepositions from Zwarts (1997)</td>
<td>127</td>
</tr>
</tbody>
</table>
Appendix D: Sources of the 38 Million Word Corpus ............................................ 130

Bibliography ................................................................................................................... 134
List of Tables

Table 1 ................................................................................................................. 1
Table 2 ................................................................................................................. 13
Table 3 ................................................................................................................. 15
Table 4 ................................................................................................................. 23
Table 5 ................................................................................................................. 30
Table 6 ................................................................................................................. 31
Table 7 ................................................................................................................. 33
Table 8 ................................................................................................................. 34
Table 9 ................................................................................................................. 35
Table 10 ............................................................................................................... 36
Table 11 ............................................................................................................... 47
Table 12 ............................................................................................................... 49
Table 13 ............................................................................................................... 54
Table 14 ............................................................................................................... 59
Table 15 ............................................................................................................... 68
Table 16: WordNet Semantic Category Ranking (Based on Percentage of Total)....... 73
Table 17 ............................................................................................................... 82
Table 18 ............................................................................................................... 82
List of Figures

Figure 1: Percentage of $p_{log10}$ values in split construction .............................................................. 37
Figure 2: Percentage of $p_{log10}$ values in unsplit construction .............................................................. 38
Figure 3: All verbs in the split form with positive $p_{log10}$ values ............................................................ 46
Figure 4: All verbs in the unsplit form with positive $p_{log10}$ values ........................................................ 48
Figure 5: $p_{log10}$ values (x) vs. frequency (y) in split construction ............................................................. 51
Figure 6: $p_{log10}$ values (x) vs. frequency (y) in unsplit construction ........................................................ 52
Figure 7: Percentage of semantic categories in split construction ............................................................... 61
Figure 8: Percentage of semantic categories in unsplit construction .......................................................... 65
Figure 9: Pct. of pos. & neg. values for transitivity per semantic category: split ....................................... 70
Figure 10: Pct. of pos. & neg. values for transitivity per semantic category: unsplit ............................... 71
Figure 11: Pct. of pos. & neg. values for transitivity per semantic category: split & unsplit ............... 72
Figure 12: Bolinger's linear modification ..................................................................................................... 75
Acknowledgments

I feel that I should write my thanks to my advisor, Prof. Robert Kirsner, in Dutch and Afrikaans. Were it not for his undying passion for these languages and unhinged linguistic curiosity, I may not have found this path. Throughout these years, he has stood by me, answered my questions, questioned my answers, shared many hours with me in discussion of linguistic minutiae, and guided me toward the end goal. I feel extremely lucky to have been able to work with Prof. Kirsner and I am also sad that his retirement will mean fewer will have the chance to learn from him as I did. Bedankt voor alles, ik zal U nooit vergeten!

To my other committee members, I also owe my gratitude. Prof. Christopher Stevens provided me with years of reading material, even if most of it comes from Scripture. During his courses on older Germanic languages, I thoroughly enjoyed myself parsing, translating, and discussing the finer intricacies of Frisian law. I have also benefitted greatly from his help and guidance in the workplace. Through Prof. Stevens, I was able to continue doing what I love most: teaching German and sharing these older languages with students. Prof. Hongyin Tao has shown me new methods for working with languages. His course on corpus linguistics and our subsequent discussions about my dissertation have pushed me to find more ways to integrate Big Data into linguistics. Hence, I am now working closer with statistics, programming, and natural language processing. These fields continue to contribute to my further development as a linguist. And although not on my committee, I would also like to extend my thanks to Prof. Olga Yokoyama. It was in her course on functional linguistics that I was first able to work on this topic, albeit observing from German and not Dutch. Her comments and questions on my final paper helped me to shape this dissertation.
I would also like to thank the Nederlandse Taalunie for their support. I was able to participate in an intensive Dutch language course in the summer of 2013, which ultimately led to my love for Vlaams, kriek, Bart Peeters, and *de zachte G*.

To Michael Anthony Ibarra I am indebted. I was able to make sense of Los Angeles via Killsonic. The weekly rehearsals, the performances, the performers all gave me the opportunity to find and to lose myself in my passion. Readjusting to life in the US after being away was difficult, but being surrounded by amazing musicians, and under the guidance of Mr. Ibarra, I was able to ground myself on the unstable tectonic plates that shifted under our feet wherever we played.

A world of thanks to Mariam Kaddoura. You've been there through almost all of this, and you've listened to me speak about my dissertation—even if you didn't initially understand some of it. Outside of me and my committee, you probably know more about this dissertation than anyone else. You've been a huge encouragement and comfort to me throughout these years. I hope to one day help you as you have helped me.
Vita

EDUCATION

MA: German Language and Culture
Arizona State University, Tempe, AZ
December 2006

BA: German Language and Culture
Arizona State University, Tempe, AZ
August 2004

German as Second Language (Exchange Year)
Universität Heidelberg
AY 2003-2004

PUBLISHED MANUSCRIPTS


CONFERENCE PRESENTATIONS

January 19-21, 2016: Conference for Young Scholars:
"Dutch Pronominal Adverbs: Meaning-Full or Meaning-Less?"
University of Stellenbosch, South Africa

May 8-9 2015: Germanic Linguistics Annual Conference:
"Rent is only die halbe Miete"
BYU, Provo UT
OTHER CONTRIBUTIONS

- Co-Editor- New German Review (UCLA Germanic Languages Grad Student Journal): 2010-2014
- Housing Committee- AMPRA 2014 Conference: January-October 2014
- Humanities Council- Departmental Representative: 2012-2014
- Graduate Student Association Forum Representative for Humanities Council: 2013-2014
- Volunteer- UCLA CompLit Polyglot Mixers: 2014
1 Problem

Pure synonymy does not exist. This is applicable to both the lexicon and the grammar. The following dissertation asks how apparently "synonymous" grammatical structures can vary in meaning and use. In Dutch, the two forms below are widely accepted as synonymous (Aarts & Wekker, 1987) (Haeseryn, 1997), (Klooster, 2001):

(1) *Daar heb ik een probleem mee.* [split]

*there have I a problem with*

(2) *Daarmee heb ik een probleem.* [unsplit]

*therewith have I a problem*

'I have a problem with that.'

A quick search in Google shows that there are far more split than unsplit forms when it comes to the above sentence:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Split form</td>
<td>36,000 (approx.)</td>
</tr>
<tr>
<td>Unsplit form</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 1

What causes this discrepancy? Why are there roughly 5000 times more split versions of the same expression than unsplit?

The item under investigation is a particular phenomenon in Modern Standard Dutch whereby pronominal adverbs (similar to English *therewith* and *hereby*) may appear in two forms, one with its two elements fused (as in *daarmee*) and one in which the two parts are separated (as in *daar...mee*). An English counterpart would be *This is the house wherein I live vs. This is the house I live in.* In nearly all handbooks and academic articles on the subject, it is assumed that the two forms, split and unsplit, are synonymous. The goal of the present study is to test the
assumption that they are synonymous and to determine what causes the choice between the two forms and whether there is a difference in meaning between the two, and if so, if there exists a noticeable difference between these two. If such a difference can be shown to exist, then the Dutch adverbial pronouns could be added to the long list of paired linguistic forms which were once claimed to be synonymous but which have been definitely shown not to be.  

Bolinger states, "a difference in syntactic form always spells a difference in meaning" (1968, p. 127). In language, when posed with different possibilities for expressing roughly the same idea, a speaker needs to determine which option best suits his/her communicative needs. The case may be that the choice depends on the situation or they may initially appear synonymous. The latter of the two options is difficult to conceive.

Another situation in which grammatical signs are claimed to be devoid of meaning is when either of two (or more) signs can occur in a given context without apparent difference in meaning. Such situations are usually described as "free variation", with the motive for the choice between the alternatives seen as being caused by stylistics or sociolinguistics. When examined from a sign based perspective, as is done in the final test, it can usually be shown that: (a) each of the contrasting meanings is appropriate for the communication of the particular messages in which they apparently alternate (b) if one goes beyond native speaker reports of "synonymy" and examines actual language use, one can find that the forms do not in fact vary freely, in semantic terms (Contini-Morava, 1995). Therefore, it would be wise to suggest in situations where it is not clear that one form may be better than the other, that one does not assume that the two forms are synonymous.

---

1 As examples: passive sentences versus active sentences in many languages, or the English pronouns *someone* versus *somebody*. For *someone* versus *somebody*, see Bolinger (1976).
In Marina Gorlach's (2009) sign-oriented work on continuous and discontinuous phrasal verbs in English, she argues that expressions in English, such as "I take off my shoes" and "I take my shoes off" are not the same. On the surface, the two forms appear similar. Her analysis states that discontinuous structures, such as "I take my shoes off", are associated with perfectivity and its counterpart, "I take off my shoes", exhibits a neutral reading, i.e. the action is either being completed or is completed. These two options, considered to "mean the same thing", indeed are not.

Elsewhere, Wulff (2006) analyses two forms also considered synonymous. Her article questions previous studies on go-Verb and go-and-Verb constructions. She argues that the second form cannot be, as stated in previous studies, truncated surface forms (Wulff, 2006, p. 103). Wulff bases her study on Adele Goldberg's Principle of No Synonymy: If two constructions are syntactically distinct, they must be semantically or pragmatically distinct (Goldberg, 1995). Wanting to determine the difference in meaning between the two forms, Wulff approaches this question from using Construction Grammar (a theory discussed later below) and bases her results on corpus data. Should the two forms be synonymous, then the constructions could schematically be used with any verb.

Wulff shows particular verbs collocate with one of the two constructions and that they are not as synonymous as assumed. She determines this by means of three statistical methods. Each of these methods aids in determining the collocational strength of different verbs with the construction. Wulff discovers that the verbs that collocate with the go-Verb construction are often atelic, expressing only the beginning of an action, and that the go-and-Verb constructions are telic, embodying the entire action of the verb. She thus exhibits yet another case where
synonymy has been suggested, but through corpus data and statistical tests, such as
collostructional analysis, the results show a different picture.

As these studies suggest, different meanings should be able to be found in variation,
otherwise variation is superfluous.\(^2\) This study suggests the same. If the form is split, then I
suggest a continuation of the subject matter before the pronominal adverb, and if it is unsplit
there tends to be a shift or narrowing of the subject matter. Bolinger's (1952) concept of Linear
Modification would suggest a narrowing of subject matter with the unsplit pronominal adverb.

Bolinger's model states that speakers often begin statements by narrowing the field of importance
and arrive at the main point by the end of the statement. By accepting this as an abstract
realization of a function of the unsplit forms, then one can presume that any information after the
construction will hone in on the subject matter or possibly signal a shift.

1.1 Object of Inquiry

As a very basic definition, the pronominal adverb in Dutch consists of two parts: the first
element (the adverb) and the second (the preposition). A more detailed description is provided in
the ANS (\textit{Algemene Nederlandse Spraakkunst}) and it is this description that will be used to
define pronominal adverbs. The ANS is the current definitive source for Dutch grammar and
usage. The two-volume guide is to be used in a similar fashion to Duden's \textit{Grammatik der
deutschen Gegenwartssprache} as well as Quirk et al. \textit{A Comprehensive Grammar of the English
Language}. It is meant both for the native speaker of Dutch and learners (Haeseryn, 1997). The
ANS defines pronominal adverbs thus:

The first element in a pronominal adverb is one of the following adverbs:

\(^2\) One question asked has been is whether the elements of a pronominal adverb, when split, should be seen as
discontinuous constituents or as two separate entities (cf. Klein 1987). For the purpose of this study, I identify split
pronominal adverbs as discontinuous constituents.
1. Demonstrative Adverbs
   a. **hier** 'here'
   b. **daar** 'there' (emphatic form of *er*)
   c. **er** 'there'

2. Interrogative Adverbs
   a. **waar** 'where/what'

3. Relative Adverbs
   a. **waar** 'where'

*Waar* is seen in questions and in relative clauses and will not be discussed in this paper. It will be explained later why this adverb is not included. The second element is a preposition, whereby it must also be noted that not all prepositions are equal. The ANS mentions that certain prepositions cannot combine with this construction, but no compelling reason is given outside of "it concerns mostly prepositions one uses in formal or technical language" (Haeseryn, 1997, p. 493).³ An example of such a preposition is *niettegenstaande* "notwithstanding".

Zwarts (1997), not content with the reasoning provided in the ANS, separates the list of prepositions from the ANS into two groups. The "A group", syntactically simple and often considered "inherited" prepositions, are those which combine with the adverbs. They are often monosyllabic and can be attested over long periods of time.⁴ The "B group" cannot combine with the adverbs.⁵ Those belonging to the B group are often prepositions that consist of adverbs and nouns that underwent semantic/grammatical change and became prepositions. On account of

³ Het betreft meestal voorzetsels uit de formele taal of uit vaktaal.

⁴ Claudio di Meola (2000) has written a monograph on the grammaticalization of German prepositions and Wich-Reif (2008) on the history of German prepositions. As for Dutch, there is a dearth of ample documentation of older forms of written Dutch to properly account for significant shifts over hundreds of years.

⁵ This has, however, been disputed in (van der Horst, 2008).
the B group's inability to form pronominal adverbs they take a different form, namely: preposition followed by a pronoun. Often when referring to a person, this construction is also standard. A full list of these prepositions is available in Appendix C.

1.2 Function of Pronominal Adverbs

The functions of pronominal adverbs, as per the ANS, are as follows (Haeseryn, 1997, pp. 494-495):

1. as an adverbial adjunct
   a. Waar denk je mee te rijden?
      what think you with to drive?
      'What are thinking of driving with?'
   b. Daarom schiet ik.
      that.around shoot I.
      'I shoot around that.'
   c. Leg het daar maar neer.
      lay it there but down
      'Just lay it down over there.'

2. as a relative pronoun

   Het huis waarin ik gewoond heb, word vernietigt.
   the house where.in I lived have, was destroyed.
   'The house (that) I lived in was destroyed.'

3. as a prepositional object

   Daar had hij echt niet op gerekend.
   there had he truly not on reckoned

---

This can also be understood as: 'Because of that, I shoot.'
'He truly did not reckon with that'/He really didn't take that into account'

4. as a nominal adjunct

Het resultaat daarvan was dat de moeder niet naar huis kwam.

the result there.of was that the mother not to home came

'The result of that was that the mother did not come home.'

It is important to note that although words like daarom 'therefore' and waarom 'why' appear to have the form of a pronominal adverb, in fact, have different functions. Taking this into account, it was necessary to choose a pronominal adverb that could not be seen as a separate lexical item. It must be able to be split and unsplit without causing confusion.

1.3 Choice of Pronominal Adverb: Daarmee

This study restricts itself to one adverb, namely daar and one preposition, met 'with'. The choice for these two units is based on:

1. The propensity of daar to be in both split and unsplit forms as er, the shortened form of daar, is less likely to be unsplit

2. Of the many prepositions, met appears to be one with some of the strongest polysemy

Met can have multiple uses.

1. instrumental:

Met de investering zijn vele tientallen miljoenen guldens gemoeid.

'Multiple millions of Guilder are at stake with the investment.'

2. comitative:

\footnotesize{7 Words such as daarom 'therefore', waarom 'why', waarbij 'whereas', daarmede 'hence' have fossilized in their form and adopted a new adverbial function. These composites cannot be split as the meaning they now have stands in contrast to what other similar constructions may signify. These particular words are pronominal adverbs on the surface, and can be used as such in some cases, but their adverbial use outweighs their use as a pronominal adverb.}
In overleg met de dokter of het Consultatiebureau voor Zuigelingen wordt vrijwel altijd een voeding gevonden waar baby goed op reageert.

'In consultation with the doctor or with the office of infant health there is almost always a diet a baby reacts well to.'

3. ornative (endowed with):

Verzamelaar van meubels, schilderijen, porselein en antiquarische boeken, wiens kasboeken evenwel niet klopten met zijn kas.

'Collectors of furniture, paintings, porcelain and antique books, whose cash account books however do not coincide with his till.'

Further explanation for why these two were chosen is made below.

1.4 Met

Met is intriguing not solely on account of its inherent polysemy, but also because its antonym, zonder 'without', does not generally occur as a pronominal adverb *daarzonder, *erzonder. The ANS mentions that pronominal adverbs, such as waarzonder 'where-without', may not be acceptable to all, yet one need not object to them (Haeseryn, 1997). Further explanation of why it is not acceptable is not given.

In Dutch, since splitting has become part of the standard language, it is also possible to abandon the adverbs daarler/waar with pronominal adverbs:

(3) Maar met geocaching kon ie [er] niets mee beginnen.8

but with geocaching could he [there] nothing with begin

'Yet he couldn't do anything with geocaching.'

---

(4) Ik heb wel internet, maar kan [er] niets mee downloaden.\(^9\)

'I have indeed internet but can [there] nothing with download

'I've got internet and all, but I still can't download anything with it'

(5) Maar.. maar ik heb [daar] geen geld mee"\(^{10}\)

but but I have [there] no money with

'But...but I don't have any cash on me'

Why this occurs is not going to be the focus of this study, but it is intriguing to note how *met* can be an orphaned preposition (Fleischer, 2000). Nevertheless, occurrences of this phenomenon are difficult to find for Dutch and especially complicated for German due to a lack of dialect corpora.

1.5 *Daar*

As seen below, there are a few adverbs that can form with prepositions to create pronominal adverbs. Two of which that are not included in this study are the less emphatic form of *daar*, namely *er* (demonstrative), and *waar* (interrogative). The splitting of pronominal adverbs with *er* tends to occur more often, thus this study avoids them as there are not enough unsplit forms for a balanced comparison. The interrogative form *waar* has also been left out for two reasons. The first reason pertains to its location in a clause. Often, *waar* appears in constructions, functioning as a relative pronoun:

(6) het mes waarmee ik het brood snij

the knife which-with I the bread cut

'the knife with which I cut the bread'

---

\(^9\) Source: http://www.androidworld.nl/forum/asus-transformer-pad-tf300t/44798-geen-wifi.html

\(^{10}\) Source: http://forums.marokko.nl/showthread.php?t=4537596&s=cc3cc71406f14f166b8e2ddb5582001b
When occupying the role of a relative pronoun, it places the verb in final position. The proximity to the verb, both syntagmatically and semantically, is one of van der Horst's (1992) reasons for why splitting occurs.\textsuperscript{11} Since his analysis makes prolific use of these types of constructions it would superficially appear that the semantic relation of verb to the preposition could be the main reason for splitting. Nearly half of the 37 sentences show a split form with a relative pronoun. Also, in van der Horst (1991), he discusses how the \textit{waar} form has overtaken \textit{daar} in relative pronouns with prepositions from the 14th to the 20th centuries. The overabundance of \textit{waar} forms as relative pronouns begs the question as to whether the examples were carefully selected. Therefore, the use of \textit{waar} in such sentences remains problematic due, in part, to how it forces verb-final structures.

Another reason \textit{waar} is excluded from this study is one of practicality. The sheer number of tokens can be overwhelming. For \textit{daarmee} alone, the entire corpus provides 6766 results and \textit{waarmee} appears 3912 times. Whereas it may very helpful to have all of these instances (10678 results in all), it can also hinder progress. If patterns can be found with one form, such as \textit{daarmee}, then looking for similarities with the other pronominal adverbs might be easier with a laid out and practiced plan of action with clearly set parameters.

\textsuperscript{11} Van der Horst's analysis will be mentioned later. Suffice it to say that, although inventive, his analysis falls short of describing why the splitting occurs beyond a possible semantic connection to the verb.
1.6 Summary of Study

The goal of this dissertation is to observe the aforementioned structure in Dutch and provide both a motivation for the variation as well as a difference in meaning. The pronominal adverb in focus is *daarmee* "therewith", and the approach is corpus-based. Considering the large shift toward more data-driven studies in linguistics (Levinson & Evans, 2010), it would appear that more empiricism and less intuition would be the order of the day. As such, this study utilizes the 38-million-word corpus of the Institute of Netherlandic Lexicology. In addition to challenging the purported synonymy of the two forms, the dissertation also aims at contributing to the growing number of studies utilizing corpus-based, empirical approaches. It shall also demonstrate how these approaches can be integrated with literature. The final test utilizes the hypotheses suggested during the first three tests and seeks a qualitative meaning difference.
2 Construction Grammar

One of the main theoretical approaches used in this study is Construction Grammar (CxG), which may be considered a particular kind of Cognitive Linguistics, cf. introductions on the topic by (Croft & Cruse, 2004) and (Langacker, 2008). It is to be noted that there are multiple approaches which label themselves as Constructionist (Boas & Sag, 2012, p. 2). For the purposes of this work, the general underlying tenets of CxG are important and, as of now, no particular approach to CxG has shown itself to be more advantageous over another. Still, the underlying premise of Constructionist approaches bears significance in this study: two constructions cannot be synonymous (Goldberg, 1995), (Croft, 2001) (Haiman, 1980).

To begin with, CxG does not take a componential approach to language; syntax and lexicon work together and form a continuum. This continuum contains multiple units which are called constructions. Constructions are conventionalized form/meaning pairing (Goldberg, 1995). A construction can be defined further as:

Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency (Goldberg, 2006, p. 5).

The predictability of these constructions depends on their nature: a construction can be anything from a morpheme to a complex construction, such as the passive. The table below, adapted from (Goldberg, 1995, p. 220), illustrates the different levels of constructions:
For example, partially filled idiomatic constructions such as "Math is a pain in my rear" [X BE pain in my Y] are less schematic than the resultative construction "X causes Y to become Z" [X V Y Z] "Steve wipes the table clean". In the first construction, only two lexical items can be placed in the slots of X and Y, and there are restrictions on what can be inserted. X represents the subject and Y is the location where X is a metaphoric pain. In the second construction, the verb (V) is open to generally any transitive verb. X can be any subject and Y any object. Z is an adjective. Hence the construction schemata allow for sentences such as: "Smith washes the car spotless", "The priest preaches the congregation asleep", and "My dog howls the neighbor insane". Each of these sentences has the underlying resultative construction [X V Y Z]. The network of these numerous constructions in a speaker’s mind is known as a construction- similar to a lexicon.

For this study, the data collected through the 38-million-word corpus provide the foundation upon which the two constructions are posited. Since I believe that two different
syntactic structures signify two different meanings, there is a need to suggest individual constructions, a suggestion which is not problematic. What this does suggest, however, is that these two constructions have two separate meanings. For CxG, each construction has its own meaning and thus these two can be schematically represented as:

A) $[[DAAR] [Preposition]] [X]$

B) $[[DAAR] [X] [Preposition]]$

X represents the remaining content in the clause. These structures are similar to the abstract models put forth by Honselaar (1994), yet differ from his as they represent the surface forms and not the abstract forms he suggests (to be discussed in the conclusion). For construction A, representative of the unsplit form, the information coming after the pronominal adverb is separated. It does not form a coherent unit with the pronominal adverb. It is possible that information in X may be similar to that which preceded the pronominal adverb, but not necessarily. Construction B, representative of the split form, represents a form where information in X is pertinent to the pronominal adverb; it is sandwiched in between. Therefore, two constructions bear two different meanings for the information in X. Information before the pronominal adverb in construction A is less likely to be related to X, whereas the information in X is more likely to be related in construction B.
3 Source Material (Corpus discussion)

The data for this study are from the 38-million-word corpus of the Institute for
Netherlandic Lexicology. All source material in the corpus derive from Standard Dutch (AN: 
*Algemeen Nederlands*) sources, printed in the Netherlands, not Belgium. The corpus, therefore,
does not allow for study of cross-border use, thus eliminating a discussion of
Flemish/Netherlandic Dutch variation. Limiting the sources to only the Netherlandic side of the
Dutch language continuum is beneficial. By only containing one variant of Dutch, the corpus
allows for a much easier reading of the results without the need to account for dialect variables.

Material in the 38-million-word corpus is a collection of printed material and composed
of 6 genres. Below is a brief list of the genres. A full list of the sources is available in Appendix
D.

<table>
<thead>
<tr>
<th>Genre</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative</td>
<td>14 books</td>
</tr>
<tr>
<td>Medicine</td>
<td></td>
</tr>
<tr>
<td>Linguistics</td>
<td>1 book</td>
</tr>
<tr>
<td></td>
<td>1 journal</td>
</tr>
<tr>
<td>Popular Science</td>
<td>2 journals</td>
</tr>
<tr>
<td>Environment</td>
<td>3 journals</td>
</tr>
<tr>
<td>Politics</td>
<td>2 journals</td>
</tr>
<tr>
<td></td>
<td>1 law gazette</td>
</tr>
<tr>
<td></td>
<td>1 course</td>
</tr>
<tr>
<td>Crown Speeches</td>
<td></td>
</tr>
<tr>
<td>Sports</td>
<td>1 newspaper</td>
</tr>
<tr>
<td></td>
<td>1 journal</td>
</tr>
<tr>
<td>Law</td>
<td>1 corpus of legal texts</td>
</tr>
</tbody>
</table>

Table 3

The final source, the corpus of legal texts, was omitted from this study due to the unnatural
nature of legalese. Source texts, however, do not play a major role in the first three tests of this
dissertation, save for in test three, where examples from the same source (i.e. same issue, book,
date) are removed. Otherwise, little attention is paid to the source of each extract. As this dissertation aims to find a general tendency in Standard Dutch, narrowing the extracts for tests 1-3 to particular sources would impede the results.

In finding the extracts for the study, results were somewhat easy to pull for the unsplit form, as the corpus is able to recognize it as a unit. With the split form, however, it was more difficult. We programmed the corpus to return results with *daar* followed by 1-7 words then *mee*. This initially provided numerous results, and these had to be checked, as the results delivered many cases such as the following, where the *mee* sought is not attached to the *daar*, but belongs to an entirely different clause. This is seen in the following examples:

(8) Maar er zijn er ook veel die het niet redden. Duizenden kinderen trekken naar de stad, in de hoop dat ze *daar* hulp krijgen. Maar dat valt niet *mee*.

But there are there also many who it not save. Dozens children travel to the city, in the hope that they *there* help get. But that falls not *with*.

'But there are many who do not make it. Dozens of children travel to city, in the hopes that they will receive help there. But that's very difficult.'

(9) Het huidige tijdsgewricht vraagt op kwantificeringen en je kunt je *daar* niet helemal van distantiëren. Zie hadden *mee* moeten denken en beslissen en dat hebben ze niet gewild.
The present era asks on quantifications and you can yourself **there** not at all from distance. They had **with** must think and decide and that have they not wanted.

'The present era expects quantifications and you cannot distance yourself from that. They should have contributed and decided, and they didn't want that.'

The following section contains the four tests. Tests one through three are thoroughly quantitative in nature. Tests one and two do not make any attempt at suggesting a meaning difference, but rather a motivation for why a splitting occurs. In test three, I use the data in from the previous tests to suggest a possible meaning difference. This meaning difference is solidified in the final test, where I observe how the meaning variation functions in extended passages of literature.
4 Introduction to the Tests

Four tests have been created to search for both the motivation and meaning in the split and unsplit forms of *daarmee*. These tests have been planned so that the dissertation focuses first on a micro-level analysis, on the particular phrasal verbs that take *met* as their preposition. This test questions a theory purported by van der Horst (1992), stating that splitting occurs on account of the cohesion of the verb to the pronominal adverb. Test two shifts the focus from the phrasal verbs in test one and their attraction or repulsion to the pronominal adverb, and it uses the frequency of *all* verbs occurring inside of the target constructions. The second test suggests a correlation between the verb's individual frequency of use and tendency to occur in either split or unsplit forms. Following the two data driven tests, the third test maintains a data driven approach, but also asks the question of what *kind* of verbs occur inside of these constructions. Answering this test helps to conceptualize a difference in meaning between the two forms. Test three investigates the semantic categories to which the verbs belong, and how these semantic categories bear on the form-meaning pairing suggested above. Two works from Dutch author Willem Fredrik Hermans, *Nooit meer slapen* and *De donkere kamer van Damocles* form the corpus for the final test. For this final test, the point of departure is no longer the single pronominal adverb, *daarmee*, but rather all pronominal adverbs paired with *daar*. During test four, I offer a demonstration of how the hypotheses put forth in the first three tests can be used in conjunction with literature to provide a deeper understanding of how motivation of use and meaning difference can be better understood. Yishai Tobin’s concept of Semantic Integrality lays the theoretical foundation for this final test.
4.1 Previous Studies into Split Pronominal Adverbs

"Those who work with language know that there is no such thing as a true synonym" (Rodale, 1986). Should there be no synonyms, then why does a consultation with the ANS (standard grammar of Dutch) only provide a succinct answer:

"Deze splitsing brengt gewoonlijk geen betekenisverschil met zich mee, maar kan als een stilistische variant beschouwd worden" (Haeseryn, 1997)

This splitting does not normally preclude any particular change in meaning, but it can be seen as a stylistic variant [Translation mine].

The ANS goes on to proclaim that the two forms vary based on location – more splitting in The Netherlands than in Belgium – and also based on medium – spoken tends to split more than written. In other grammars, the story remains the same: no difference in use beyond the spoken/written dichotomy. A survey of various Dutch grammars has returned unexpected, but also disappointing results. One would hope that definitive grammar sources would provide insight into a problematic phenomenon, particularly for learners, yet one remains disappointed.

Some grammars either ignore the subject of the split and unsplit pronominal adverbs entirely (Rijpma & Schuringa, 1969), and others provide only a brief statement to the effect that " [...] the separated forms are the usual ones in spoken Dutch" (Kruisinga, 1949). Kruisinga does not provide any further information. One grammar indicates that the splitting is an "odd" phenomenon: "Vreemd genoeg kunnen de woorddelen wel gescheiden worden" (Houët, 1988, p. 60). Why this would appear odd does not warrant an explanation. Van den Toorn (1984, p. 233) states that there is no recognizable difference in meaning between the two forms: "De leden worden gewoonlijk om stilistische redenen gescheiden, zonder dat men verschil in betekenis kan

---

12 Odd enough, the parts can be split. [Translation mine]
opmerken, bijv. *Ik ben daarover niet teverden of Ik ben daar niet tevreden over*. In only one grammar, (Donaldson, 2008), does there appear any slight semblance of meaning difference. In his grammar of Dutch for English speakers, Donaldson suggests that the split form emphasizes the anaphoric reference.

\[
\begin{align*}
(10) & \quad \text{Ik heb het daarmee (= daar + met) geschreven} \\
& \text{I have it there.with written} \\
& \text{‘I wrote it therewith’}
\end{align*}
\]

would be the standard, non-emphatic use. In contrast, the split form emphasizes the object with which was written.

\[
\begin{align*}
(11) & \quad \text{Daar heb ik het mee geschreven} \\
& \text{there have I it with written} \\
& \text{I wrote it with THAT. (emphasized).}
\end{align*}
\]

More information regarding how one could justify this emphatic reading does not appear. One would expect Donaldson to support his reasoning for marking the split form as one of emphasis, yet he abandons the topic forthwith.

Beyond grammars, Dutch style guides also do not make much of an effort to explain this dichotomy. They do underline that there are particular words that appear to be pronominal adverbs, such as *waarom, daarom, daarop, daarnaast*, and *waaronder* that cannot be separated, as mentioned above. These words now already have their own meaning (Klein & Visscher, 1996). The abovementioned words can be split should that "more or less idiomatic meaning" is no longer there. In another style guide, the reader is warned of the temptation to split the forms. Instead, one is to write them together, as it is correct (Houët, 2000).

---

13 The parts can be split for stylistic reasons without that one would notice a change in meaning. For example, "I'm not happy about that".
Only a few linguists have taken the standard grammars to task. An earlier analysis of the split and unsplit forms can be found in Van Riemsdijk (1978). In his paper, van Riemsdijk argues in the Generativist mode. His analysis rests upon mechanical syntax extraction and not on motivation or semantic variation in meaning. For him, "the more closely a prepositional phrase is connected with the verb, the easier it is to extract elements from such a prepositional phrase" (van Riemsdijk, 1978, p. 26). It remains unclear as to how he defines "closely [...] connected". His is pure syntactic movement, no offer of meaning difference or deeper analysis of these two forms is made. Thus, the reader is left still wondering: Why? If it is only mechanical and there is no need to discuss any semantic variation, for what purpose do speakers of Dutch choose to use one form over the other?

Joop van der Horst (1992) builds on van Riemsdijk (1978) and states that a reason for the pronominal adverb split is connected with "the degree of cohesion between the adverbial pronoun and the verb" (van der Horst, 1992, p. 127). His analysis is quite similar to that of his van Riemsdijk. Where van der Horst differs, however, is in his detailed breakdown of when he believes the form can split and when it cannot. In essence, he postulates some minimal motivation behind the split form. The table below represents his breakdown. We have added a third column to provide a summary of his reasoning (van der Horst, 1992, pp. 134-135).

<table>
<thead>
<tr>
<th>Splitting Needed</th>
<th>Dutch</th>
<th>English</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Je moet zeggen</td>
<td>You've got to give it to me</td>
<td>Fossilized idiomatic</td>
<td></td>
</tr>
<tr>
<td>waar het op</td>
<td>straight</td>
<td>expression</td>
<td></td>
</tr>
<tr>
<td>staat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dat is iets</td>
<td>That's something I demand</td>
<td>Fossilized idiomatic</td>
<td></td>
</tr>
<tr>
<td>waar ik op</td>
<td></td>
<td>expression</td>
<td></td>
</tr>
<tr>
<td>sta</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch expression</td>
<td>English translation</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>---------------------</td>
<td>--------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Daar komt het op neer</td>
<td>That's the point</td>
<td>Fossilized idiomatic expression</td>
<td></td>
</tr>
<tr>
<td>Daar heb ik lang naar uitgezien</td>
<td>I was looking forward to that for a while</td>
<td>Intuition</td>
<td></td>
</tr>
<tr>
<td>Daar kan je op rekenen</td>
<td>You can count on that</td>
<td>Intuition</td>
<td></td>
</tr>
<tr>
<td>een boek waar veel mensen van gehoord hebben</td>
<td>a book many people have heard about</td>
<td>Intuition</td>
<td></td>
</tr>
<tr>
<td>Daar werden we nogal door afgelied</td>
<td>We were quite diverted from that</td>
<td>Intuition</td>
<td></td>
</tr>
</tbody>
</table>

### Splitting Desired

<table>
<thead>
<tr>
<th>Dutch expression</th>
<th>English translation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daar heb ik lang naar uitgezien</td>
<td>I was looking forward to that for a while</td>
<td>Intuition</td>
</tr>
<tr>
<td>Daar kan je op rekenen</td>
<td>You can count on that</td>
<td>Intuition</td>
</tr>
<tr>
<td>een boek waar veel mensen van gehoord hebben</td>
<td>a book many people have heard about</td>
<td>Intuition</td>
</tr>
<tr>
<td>Daar werden we nogal door afgelied</td>
<td>We were quite diverted from that</td>
<td>Intuition</td>
</tr>
</tbody>
</table>

### Splitting and Non-Splitting Equally Usable

<table>
<thead>
<tr>
<th>Dutch expression</th>
<th>English translation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>het mes waarmee ik snij</td>
<td>the knife with which I cut</td>
<td>No difference in meaning</td>
</tr>
<tr>
<td>het mes waar ik mee snij</td>
<td>the knife I cut with</td>
<td>No difference in meaning</td>
</tr>
<tr>
<td>Daarvoor zijn we nog in Keulen geweest</td>
<td>We were in Cologne for that reason</td>
<td>No difference in meaning</td>
</tr>
<tr>
<td>Daar zijn we nog voor in Keulen geweest</td>
<td>We were in Cologne for that reason</td>
<td>No difference in meaning</td>
</tr>
</tbody>
</table>

### Splitting Less Desired

<table>
<thead>
<tr>
<th>Dutch expression</th>
<th>English translation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>de manier waarop hij werkt</td>
<td>the manner in which he works</td>
<td>Concrete location</td>
</tr>
<tr>
<td>het kamertije waarin hij zijn</td>
<td>the room in which he wrote</td>
<td>Concrete location</td>
</tr>
</tbody>
</table>
Laatste roman schreef

HIS LAST NOVEL

de snelheid waarmee hij schrijft

THE SPEED WITH WHICH HE WRITES

None given

de stijl waarin hij componeert

THE STYLE IN WHICH HE COMPOSES

None given

### Table 4

<table>
<thead>
<tr>
<th>Dutch Expression</th>
<th>English Translation</th>
<th>Fossilized Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daarvoor zijn we nog in Keulen geweest</td>
<td>We were in Cologne beforehand</td>
<td>Fossilized time expression</td>
</tr>
<tr>
<td>Waarom ga je niet met ons mee?</td>
<td>Why don't you go along with us?</td>
<td>Fossilized question word</td>
</tr>
<tr>
<td>Daardoor heb ik nog niet kunnen eten</td>
<td>Because of that I was not able to eat</td>
<td>Fossilized expression</td>
</tr>
</tbody>
</table>

Yet despite this motivation, he also maintains that the split/unsplit pronominal adverbs are "geen taalteken"; they are not a linguistic symbol (van der Horst, 1992, p. 142). Whereas van der Horst’s heuristic for understanding the expected and unexpected splitting, his article lacks

1) a clear statement regarding motivation of use between split and unsplit beyond cohesion

2) no real quantifiable evidence for this statement

---

14 Van der Horst does not provide any metrics for measuring this cohesion.
Save for a few numbers regarding split vs. unsplit forms in the written and spoken language taken from the *Frequentiewoordenboek* (van der Horst, 1992, p. 138), he fails to deliver on point two.

A third article by Wim Honselaar (1994) addresses the issue of split/unsplit pronominal adverbs and appears two years after van der Horst. Here, Honselaar (1994) argues that the difference between the two forms lies not in the cohesion between the verb and the pronominal adverb, but rather in their respective information structures. When two variants of the split and unsplit form are possible, as with the example sentence: Daar heb ik een proleem mee, Honselaar states that "splitsing is dan ook geschikter voor een evaluatieve reactie op het totaal van iemands woorden" (Honselaar, 1994, p. 35). Context matters. The split form provides a synthetic information structure (X-Y), whereby the verb and the pronominal adverb create one whole information block (Honselaar, 1994, p. 39). As such, the statement

(12) het podium *waar ik op sta*

the podium where I at stand

' the podium I stand at'

speaks more, according to Honselaar, to the continuous information flow between the act of standing at the podium and the podium itself. This form keeps the reader/listener "inside" the realm of the discussion topic of the podium. The unsplit variant,

(13) het podium *waarop ik sta*

the podium whereat I stand

' the podium at which I stand'

---

15 "Splitting is therefore a more suitable choice for an evaluative reaction to the whole of someone's words" [translation: de Carlo].
portrays an analytic information structure \((X)-(Y)\). In this case, the speaker/writer intends on focusing on the podium, so that he/she may provide further information about the podium itself. The information structure in the unsplit form keeps the listener/reader inside of the realm of the topic, but may signal something new as well.

Via Honselaar's information structure, this dissertation seeks to provide further quantitative evidence for a difference in meaning and motivation for the two forms. His synthetic/analytic duality provides a springboard for this study, but is not the foundation. In the following section, I question van Riemsdijk's and van der Horst's theory that the cohesion between the verb and pronominal adverb causes the splitting. It is also van der Horst's desire (1992, p. 143) to deliver a quantitative test of his theory, but he lacked the resources at the time.\(^{16}\) Beyond questioning their reasons for splitting, the first test also provides more background into whether or not the connection between the verb and the pronominal adverb could be a splitting factor, but not the only one.

\(^{16}\) A similar lament is expressed by Fleischer (384, 2003).
5 Test One: Phrasal Verbs

5.1 Introduction

The first test in a series of four looks solely at the attraction or repulsion to either the split or unsplit form of phrasal verbs that take *met*. An early approach to the splitting is found Overdiep in (1928, p. 281). In §398 B. he states:

"De twee woorden zijn in den zin gescheiden: vooral in de omgangstaal en ingeval het adverbium een locale bepaling is. In litterairen en ambiëtijken stijl en vooral wanneer het adverbium géén concrete, locale aanduiding is, is de neiging tot verbinding der twee woorden groot. Dit in 't algemeen. 1° De scheiding kan bovendien in de hand gewerkt worden door omvangrijken vorm van het adverbium en omstandigheden van rhythmischen aard, die moeilijk onder een regel te brengen zijn".\(^{17}\)

Were one to follow his reasoning, either the spoken form, the cadence of the utterance, a locative adverb, or an adverb without a concrete designation would cause splitting. Since, however, this study is not observing differences in spoken vs. written or the vocal cadence of utterances, these two possibilities will not be considered. This leaves locative adverbs or adverbs without a concrete designation as being culprits for the split. Van der Horst (1992) takes a footnote from Overdiep (1928, p. 610) as his motivation for looking further: "Dat het locale adv. minder gemakkelijk met het voegw. verbonden raakt, ligt voor een deel aan de nauwe verbinding van

\(^{17}\) "The two words are split in the sentence, especially in the spoken language and when the adverb is a locative one. In the literary and official styles and, in particular, when the adverb does not have a concrete, locative meaning is there a stronger tendency toward the unsplit form. This is a general tendency. The splitting can also be caused by the various forms of the adverbial and rhythmic environments, which are difficult to define" [Translation mine].
locale adv. met het werkwoord (dat aan 't einde van den zin staat)''. The verb, according to van der Horst, should be the cause for splitting.

Van Riemsdijk (1978, p. 25) also observes the splitting phenomenon. In his dissertation, he argues that "[...] prepositional objects and adverbial prepositional phrases belong to the same syntactic category, this is not to say that they behave in the same manner in all respects". One of the differences that bears on van der Horst (1992), and by extension this dissertation, suggests that adverbial prepositional phrases have different extraction possibilities. Whereas this dissertation does not use Generative Syntax, and any ideas of extraction or preposition stranding play no role, van Riemsdijk's corollary seems to go along with Overdiep (1928): "Extraction possibilities correlate with this parameter: the more closely a prepositional phrase is connected with the verb, the easier it is to extract elements from such a prepositional phrase" (van Riemsdijk, 1978, p. 26). This is therefore another case where the verb is purported to cause the split. In the next paragraph van Riemsdijk bypasses any semantic meaning variance between the two forms, maintaining the Generativist tradition at the time of separating semantics from syntax.

Building on these two, van der Horst (1992) lays out a rubric, as shown above in Table 4, describing when and where one would expect a split or unsplit form. Nevertheless, neither van Riemsdijk nor van der Horst have shown that their theories can furnish measurable results. Test one of this dissertation is purely numerical. No attempt will be made to distinguish between form and meaning, nor will any hypothesis be initially afforded that would make claims as to why

18 "The fact that locative adverb is less easily to connected to the conjunction is tied to the close connection that the locative adverb has with the verb (that is at the end of the clause)." [Translation mine]

19 No mention of Overdiep's 1928 Grammar is mentioned in van Riemsdijk's dissertation.

20 "I take this parameter to follow from constraints on extract-ability that interact with semantics and operate independently of the constraints proposed in this study" (van Riemsdijk, 1978, p. 26).
splitting occurs. The former appears in tests three and four, the latter in test two. This first test consists of three steps, outlined below. Each of these steps were crucial in questioning both van Riemsdijk (1978) and van der Horst (1992).

5.2 Problem

If one is to claim that the cause for this split/unsplit dichotomy rests in the verb, then what one necessitates is quantifiable evidence. The tools available to linguists now facilitate these tests to a much higher degree than were available before. It has been the tradition of generative linguists to create their own sentences to test grammaticality. Yet there appears to be a movement toward using corpus data to assist in finding interesting, insightful and grammatically "correct" sentences. In Meuers (2005, p. 2) he states that:

"data obtained from corpora are a highly valuable source of empirical insights which can help verify linguistic generalizations and serve as a diverse empirical basis for the development and revision of linguistic theories."

And although the creator and constant innovator of Generativism, Chomsky, may think that "corpus linguistics doesn't mean anything" (Andor, 2004, p. 97), it is difficult to ignore such comments. His comparison with elder scientists, such as Galileo, simply observing "videotapes of leaves falling and balls going around and rocks rolling down mountains and so on and so forth" (Andor, 2004, p. 24), is flawed. For if Galileo were to have videotapes at his disposal, he would surely make use of the physical data. To wholly disregard corpora as being empty ignores any contribution they can make.

With this in mind, test one sets out to show that corpus linguistics can mean something and can (and often do) deliver useful results. For this test, I make extensive use of the data collected from the 38-million-word corpus of Netherlandic Dutch. Primarily, this test is to
examine the idea that there is some connection between the verb and the split pronominal adverb. Since three academics, from three different eras and three different fields, Overdiep (1928), van Riemsdijk (1978), and van der Horst (1992), all suggest that the verb's relation to the pronominal adverb causes the splitting, it is now time to see if there is, in fact, a connection.

5.3 Methodology

After the data were properly cleaned, the first step in building this test was to determine which verbs to use. Since van der Horst (1992) directly states that the verb's attraction causes splitting, all phrasal verbs that take *met* as their preposition were sought. As the pronominal adverb I am observing is *daarmee*, this seemed to be the most logical extension. These verbs are all related to the pronominal adverb in that their secondary component, the verb *met*, rests inside. Should there be any reason to split, it may rest in the connection between these verbs and their connected preposition. The Prisma Voorzetsels dictionary (Reinsma & Hus, 2005) yields a list of Dutch verbs fitting this criterion. All phrasal verbs that take *met* as their preposition were compiled from this resource. In total, there are 346 verbs in this category.

From here follow two further steps. Step one includes the tagging of all the corpus extracts for parts of speech. This step is crucial in differentiating between verbs and other parts of speech, in particular nominalized verbs. To accomplish this task Tree-Tagger, an automated part of speech parser, tagged the parts of speech. Text preparation included proper encoding (UTF-8) and separating each individual word on a separate line. Separation on each line permitted a much easier analysis of the verbs, as it greatly simplifies the process of checking all verbs that occur with preverbs, such as *aan-knopen, af-geven, in-stemmen, weg-komen*, etc. It is often the case that tree-tagger does not register the two elements, and mislabels the verb

---

21 A full list is available in the appendix.
infinitive. Once the data were controlled for improper tagging, all verbs matching those on the Prisma list were tagged "met_verb" to aid in the counting. Below is an excerpt from the completed table:

![Table 5](image)

Table 5

---

22 Such as *af* and *geven* which form the verb *afgeven*. This occurred enough in the initial stages of test preparation, it became clear that a manual check of all entries was needed.
On the far left is the original entry, broken down word-by-word. The far right entry is the word from the original, but parsed to its bare form. For example, in line 89 beperkt is in the past participle form, but tree-tagger parses it as its infinitive, beperken. Once this is complete, tree-tagger assigns a series of numbers to each part of speech, i.e. 201 is an intransitive verb in the 1st person singular present. From these numbers, I determined which forms would be used. A list of all codes used for this test are available in the appendix. These codes were then changed from their numerical values to the tag "VERB_USE", thus making it much easier to extract and count the verbs either inside of outside of the construction.

The following step was to ensure that all verbs observed fit the criterion of being a phrasal verb that takes met as its preposition. Furthermore, the following modal and auxiliary verbs were omitted from the count:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>zijn</td>
<td>moeten</td>
<td>durven</td>
</tr>
<tr>
<td>to be</td>
<td>to have to</td>
<td>to dare</td>
</tr>
<tr>
<td>hebben</td>
<td>hoeven</td>
<td>zullen</td>
</tr>
<tr>
<td>to have</td>
<td>to need to</td>
<td>shall, will</td>
</tr>
<tr>
<td>mogen</td>
<td>kunnen</td>
<td>worden</td>
</tr>
<tr>
<td>to be allowed to</td>
<td>to be able to</td>
<td>to become</td>
</tr>
</tbody>
</table>

Table 6

These verbs were removed as they would skew the data. Omission was also necessary of the verbs zijn, hebben, and worden as they occur so frequently among the data that it would be an immense demand on time to cross-check all of them for use as full verbs vs. auxiliary verbs. Hence, they were deleted from the count and also not tested for use.23

---

23 For further discussion of some issues regarding manual counting and necessary omission, see Schmid & Küchenhoff, 2013.
The next step in preparing the data for analysis was to ensure that all verbs that were tagged with the "VERB_USE" label were properly labeled. Because the program is not 100% accurate, it was imperative to check each labeled entry. In some cases, a verb would be marked <unknown>, as can be seen in line 84 in Table 5 above, and should have been counted. Also, as mentioned above, any verbs that take a separable prefix verb had to be checked. To speed up this process, all separable prefixed verbs were reduced to their root form and cross-checked throughout the dataset. In concluding the data preparation, all verbs that belonged to the list compiled via the Prisma database were tagged with "met_verb" and the cells colored for easier recognition.

5.4 Collostructional Analysis

With the data ready for analysis, the first test begins. Test one utilizes a collostructional analysis. Collostructional analysis is a method put forth by Gries and Stefanowitsch (2003) to test the likelihood of words to be attracted to or repulsed from a particular grammatical construction. Collostructional analyses function inside of the framework of Construction Grammar (CxG). The construction schema used for this test consists of the pronominal adverb, split or unsplit, and one finite verb, as outlined above. The pronominal adverb must occur inside of a complete clause, and pronominal adverbs without finite verbs were eliminated. The exception to this deletion is in cases where the pronominal adverb appears inside of a clause with modal auxiliaries. Here, the infinitive was used in the count. Similarly, if the finite verb was one of the auxiliaries zijn, worden, or hebben, only the past participle was counted. Each construction was found manually throughout the data to ensure accurate counts.

---

24 Schmid & Küchenhoff (2013) provide a critique of collostructional analysis.

25 Further studies utilizing collostructional analysis can be found in (Gries, Hampe, & Schönefeld, 2005), (Gries, Hampe, & Schönefeld, 2010).
In order to determine attraction or repulsion, the data must first be organized so that they can provide accurate counts. These numbers go into a contingency table, or 2x2 table, illustrated below:

<table>
<thead>
<tr>
<th></th>
<th>In Construction (split or unsplit)</th>
<th>Not in Construction (split or unsplit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;met_verb&quot;</td>
<td>target lexical item in construction</td>
<td>target lexical item not in construction</td>
</tr>
<tr>
<td>not &quot;met_verb&quot;</td>
<td>non-target lexical item in Construction</td>
<td>non-target lexical item not in construction</td>
</tr>
</tbody>
</table>

Table 7

Verbs that were tagged as "met_verbs" were first counted to see how often they occur inside and outside of the constructions: split or unsplit daarmee. This was done by first finding all split and unsplit forms, coloring the cells for easier recognition, and adding labels in separate columns. This allows for a more rapid search through the data, cutting back time significantly. Following a complete count, all other non "met_verbs" (excluding those listed above) were counted inside and outside of the respective constructions.

We have adapted this collostructional analysis to test all categorized "met_verbs" for attraction. Whereas previous studies have observed the attraction or repulsion of particular lexical items to constructions (Colleman, 2009) (Stefanowitsch & Gries, 2003) (Hilpert, 2006), this test analyses a group of verbs. Collostructional analyses make use of two statistical tests: one to determine randomness of data (χ²), and one to determine attraction or repulsion (Fisher Exact). Chi square tests have been questioned in their efficacy (Davis, 2002), yet for this study, they will be utilized to determine the significance of the data. In order to move forward, it must first be determined if the data are at all significant and are random, and if so one could discard the null
hypothesis—there is no change on account of the variables. 26 The second metric used in collostructional analyses is the \( p_{\log10} \) value. In collostructional analyses, the \( p_{\log10} \) value measures attraction or repulsion to the construction in question. As a value approaches positive infinity, the construction attracts the target item, and one that moves toward negative infinity the construction repulses the target item. Values at 0 show no preference either way. In order to obtain this metric, the data must first undergo a Fisher Exact test. For this, I have used the statistical program, R, to run the Fisher Exact test.

5.5 Results

As van der Horst states "the degree of cohesion between the adverbial pronoun and the verb" (1992, p. 127) is the cause for splitting. A similar statement is made by van Riemsdijk (1978). With the database of phrasal verbs that take met as their preposition, one would expect that there would be more occurrences of these verbs among split form of daarmee, were one to accept van der Horst's and van Riemsdijk's conclusions. The numbers in Table 8 represent the count of "met_keywords" that appear in the mass corpus results for the split form and unsplit form. No consideration was taken regarding where the verbs appear, only that they appear among the mass data.

<table>
<thead>
<tr>
<th></th>
<th>split</th>
<th>unsplit</th>
</tr>
</thead>
<tbody>
<tr>
<td>met_verb</td>
<td>1169 (11%)</td>
<td>4828 (20%)</td>
</tr>
<tr>
<td>not met_verb</td>
<td>9362 (89%)</td>
<td>18898 (80%)</td>
</tr>
<tr>
<td>Total</td>
<td>10531 verbs</td>
<td>23726 verbs</td>
</tr>
</tbody>
</table>

Table 8

---

26 For this dissertation, the significance level, \( \alpha \), stands at .05.
Thus, were "degree of cohesion" the motivating factor for splitting, the numbers in Table 8 speak to the contrary. In fact, the percentage of "met_verbs" in the unsplit total is nearly twice as much. This alone may be enough to refute their claims. However, the focus of this test is not simply raw percentages, but rather the attraction of "met_verbs" to one form or the other.

Since there no preponderance of "met_verbs" occurring with the split form, as would be expected, the next step is to see just how attracted or repulsed these verbs are to their respective constructions. This section of test one analyzes how attracted or repulsed all "met_verbs" are to the two constructions. Table 9 below shows how many times a "met_verb" appears inside or outside of the split form construction. Verbs that did not belong to the "met_verb" category are also counted and placed in a separate cell.

<table>
<thead>
<tr>
<th>Split form</th>
<th>In Construction</th>
<th>Not in Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>met_verb</td>
<td>300 (26%)</td>
<td>869 (74%)</td>
</tr>
<tr>
<td>non met_verb</td>
<td>782 (19%)</td>
<td>3397 (81%)</td>
</tr>
<tr>
<td>p = 3.111 e-07</td>
<td>p_{log10} = -6.507099989</td>
<td></td>
</tr>
</tbody>
</table>

Table 9

A total of 1169 "met_verbs" occurred inside the mass data of all split form constructions (see Table 8), yet only 26% of these occurred inside of the clause with the split pronominal adverb *daarmee*. Recall that *hebben, zijn, worden* and all modal auxiliaries have all been excluded from this test due to the amount of time needed to ensure proper encoding and possible data skewing. Table 9 displays that there is not a strong affinity for phrasal verbs with *met* to occur within the split construction. In fact, the metric of attraction/repulsion for collostructional analyses, the $p_{log10}$ value, indicates that the "met_verb" grouping exhibits a repulsion, rather than an attraction, to the split form: -6.51.
For the unsplit form, the situation is similar. One noticeable difference is seen with the non "met_verbs" found inside of the construction. The percentage of non "met_verbs" inside of the unsplit daarmee is nearly 4 times as much.

<table>
<thead>
<tr>
<th>Unsplit form</th>
<th>In Construction</th>
<th>Not in Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>met_verb</td>
<td>521 (28%)</td>
<td>1358 (72%)</td>
</tr>
<tr>
<td>non met_verb</td>
<td>5270 (70%)</td>
<td>2226 (30%)</td>
</tr>
</tbody>
</table>

Table 10

The unsplit form, like the split form, also exhibits repulsion of the "met_verbs", albeit stronger. The combination of the results for split and unsplit daarmee illustrate how van der Horst's and van Riemsdijk's theories cannot stand. We may see a slight tendency for the split forms to attract slightly more than the unsplit, nevertheless, both exhibit a degree of repulsion with regards to "met_verbs".

5.6 Individual Verbs

Seeing that both forms repulse the "met_verbs", another question remains: What could be the cause of this discrepancy in repulsion? The following section will address the p_{log10} values resulting from collostructional analyses done on the individual "met_verbs" that appeared inside the constructions in the data. In order to test the individual verbs, a similar process was undertaken as above. Each verb was counted inside and outside of the construction as well as all other verbs, and a Fisher-Yates exact test was run on every verb.

Before addressing the individual verbs, let us first take a look at how many of these individual verbs provided either a zero, positive, or negative p_{log10} value. For the split forms, the Figure 1 below shows that there is a near 3-way split.
40% of the total "met_verbs" that occurred in the split form are not affected by the construction at all, 27% are attracted to the form, and 33% are repulsed. For the split form, the average $p_{\log10}$ values in the positive are 0.295 and the negative -0.743. We see here two values closer to 0, reflecting negligible change. The "met_verbs" are only slightly attracted to the split form, but are more strongly repulsed. When looking at the unsplit form, the values differ greatly.
Figure 2: Percentage of $\text{p}_{\log 10}$ values in unsplit construction

The distribution shows twice as many negative $\text{p}_{\log 10}$ values for the unsplit (66%) form than the split. The average positive $\text{p}_{\log 10}$ value for the unsplit is 0.281 and the average negative $\text{p}_{\log 10}$ values stands at -2.391. As one can see, there is a much stronger negative average value with the unsplit than the split, suggesting that something else may be influencing this difference. There is a much stronger average repulsion to the unsplit form over the split, yet this alone cannot account for the choice of either structure.27

5.7 Closing Discussion

The goal of test one was simply to uncover any possibility that van der Horst's and van Riemdijk's claims of cohesion could be supported. From the data, it has become apparent that an element beyond the verb's attraction to the pronominal adverb is causing a split. Therefore, we have a situation whereby the claims made by the two scholars cannot be leveled as the reason for

---

27 Of the 171 "met_verbs" in the unsplit form and the 100 in the split, only 43 of these verbs occur in both forms. Of these similar verbs, only four appear positive for both. Clearly there is not much overlap with regard to verb choice and either form.
splitting. Should their theories be correct, one would expect phrasal verbs that take *met* as their preposition to be attracted to the split construction. Since this is not the case, another factor must motivate the splitting. Verbs occurring in the split form should then exhibit a common trait, providing cause for the split forms. The data show the four highest ranking verbs in the split form as *omgaan* \((p_{\text{log10}}=17.61)\), *instemmen* \((p_{\text{log10}}=7.10)\), *opschieten* \((p_{\text{log10}}=4.36)\), and *omspringen* \((p_{\text{log10}}=3.66)\). In context, these appears mostly in the clauses as either idiomatic or as fossilized expressions.

*omgaan*

(14) Maar daar leren wij wel discreet mee omgaan.

but then learn we rather discreet with deal

'However, we do learn how to deal discreetly with it.'

(15) [...] dat de eilanden daar slordig mee omgaan.

[...] that the islands there carelessly with deal.

'[...] that the islands deal with that carelessly'

*instemmen*

(16) De eerste kamer heeft daar vandaag mee ingestemd.

the first chamber has there today with agree

'The senate has agreed to that today.'

(17) Daar zou geen navo-land mee instemmen.

there should no NATO country with agree.

'No NATO country would agree with that.'

*opschieten*

(18) [...] maar zijn we daar ook veel mee opgeschoten?
[...] but are we there also much with make.progress?

' [...] but are we making any progress with that?'

(19) [...] dat we daar helemaal niets mee zouden opschieten.

 [...] that we there completely nothing with should make.progress.

' [...] that we would not make any progress at all with that.'

omspringen

(20) [...] en daar nogal optimistisch mee omspringt.

 [...] and there fairly optimistic with deal.with.

' [...] and deals with it fairly optimistically.'

(21) U moet daar voorzichtig mee omspringen.

you.formal must there carefully with deal.with.

'You have to deal with that carefully.'

By contrast, the unsplit form's highest ranking verbs – aangaan ($p_{log10}=2.37$), aankomen ($p_{log10}=1.36$), aanleggen ($p_{log10}=1.12$), and aanspreken ($p_{log10}=0.40$) – all have much lower $p_{log10}$ values than their split counterparts. Yet it is striking that despite being the highest ranking verbs among the unsplit forms, only one verb occurs more than once inside the unsplit construction. In both instances, the verb aankomen appears in the following construction:

(22) Daarmee ZIJN SUBJ aangekomen bij [...]  

'To have arrived at something [...]'

Here, the writer is directing the reader's attention to arriving at a particular point in the discussion, focusing on movement forward. Contrastively the split form's two examples of aankomen ($p_{log10}= 1.32$) appear in negated forms with the modal verb moeten.
(23) Daar moet je niet mee aankomen

There must you not with arrive.

'You shouldn't have to confront that' (Bilderdijk, 1826).

(24) Daar moeten ze niet bij me aankomen

there must they not at me arrive.

'They shouldn't have to confront me with that'

In both cases, the writer anaphorically references a negative situation, and is using *aankomen* not in its literal sense of arriving (as is the case with the unsplit form), but rather in the sense of confronting someone with something. The split form appears to exhibit a predilection toward idiomatic expressions.

In sum, cohesion between the phrasal verb and the pronominal adverb can not be considered the sole cause for the split form to occur. Secondly, there appears to be a pattern with the split forms: the verbs occur in more familiar constructions, possibly idiomatic. This tendency to function as more common expressions opens a new line of inquiry: what is the relation of these expressions to the "met_verbs", and how does this relationship affect, if at all, the difference in motivation for a split or unsplit form? The following test considers how frequency could be a motivating factor in differentiating between the two forms.
6 Test Two: Frequency

6.1 Introduction

Following the results of test one, test two observes the frequency of the individual verbs. Since test one suggests that there may be a correlation between more common and frequent expressions and the split form, test two is devised as a way of furthering this point of departure. The following hypotheses direct test two:

H_0: No discernable reason exists for the differentiation in use between the two forms.

H_1: The verb's frequency will motivate use of either the split or unsplit form.

Should the test deliver results that support H_0, there may be yet another force in determining which form is chosen. If, however, H_1 shows to be true, then I would be capable of delivering a tested, quantifiable reason for the use of split versus unsplit forms. Still, no suggestion regarding the meaning difference shall be made; our goal here is to discover what could be motivating the discrepancy seen in test one.

The effect of a word's frequency can be considered a motivating factor in choice. Should the verb in the clause with the pronominal adverb be more frequent, this clause may have a lower threshold of activation, requiring less cognitive energy to process the information. This would allow the reader to skim through the text with minimal effort. In the event that the clause includes a verb lower in frequency, hence more difficult to process, then the reader must take his time in processing the clause. As an example, the indefinite articles in English, *a* and *an*, are two extremely frequent words. In DeLong, Urbach, & Kutas (2005, p. 1117), they test how quickly subjects choose *a* vs. *an* in an effort to show how the mind activates other words in a sentence to
cue the likelihood of words to follow. Their results show that the mind is quicker to predict the proper article, but also that this prediction depends on the context:

Our results thus demonstrate not only that readers can rapidly, incrementally integrate incoming words into evolving mental sentence representations, but that they do so in part by exploiting various constraining forces to form probabilistic predictions of which specific words will come next (DeLong, Urbach, & Kutas, 2005, p. 1119).

Hence, should the mind be able to predict what appears afterward, there should be less cognitive strain. The reader exercises less effort in processing the information and is able to parse what comes after with little effort. This, in effect, is what H1 states: Should the frequency of the verb be higher, then one should expect the split form to appear, as the reader may be able to predict the second element, namely the mee, before it appears. By contrast, if the clause contains a less frequent verb, one would expect the form to be unsplit, as there would be more strain on the mind to process both the verb and then to predict the second element of the pronominal adverb.

As William James states in his 1890 book, Principles of Psychology,

We all of us have this permanent consciousness of whither our thought is going. It is a feeling like any other of what thoughts are next to arise, before they have arisen. This field of view of consciousness varies very much in extent, depending largely on the degree of mental freshness or fatigue. When very fresh, our minds carry an immense horizon with them. The present image shoots its perspective far before it, irradiating in advance the regions in which lie the thoughts as yet unborn. Under ordinary conditions the halo of felt relations is much more circumscribed. And in states of extreme brain-fag the horizon is narrowed almost
to the passing word, the associative machinery, however, providing for the next word turning up in orderly sequence, until at last the tired thinker is led to some kind of a conclusion. (James, 1950, pp. 255-256).

6.2 Methodology

Test two uses Tiberius & Schoonheim's Frequency Dictionary of Dutch (2014) to determine frequency of the individual verbs. This resource includes 5000 words, subdivided into two lists and four genres. The lists and their respective highest frequency values out of 100 texts (HFV) are as follows –

- Core: Words that appear 4.5 times across the four genres. There are 943 core vocabulary words. (HFV: 99.92)
- General: words that have high frequencies across at least three genres, but are not part of the core vocabulary. 2004 words appear in this list. (HFV: 45.84)

The four genres are fiction (HFV: 8.45), newspaper (HFV: 66.8), spoken (HFV: 7.29), and web (HFV: 66.8). To determine if a word belongs in a particular genre, the dictionary makes use of four separate corpora. For fiction, 900 books from 1970-2009 are used. These books are written by both Dutch and Flemish authors. The newspaper genre utilizes the SoNaR (Stevin Nederlandstalig Referentiecorpus), extracting newspaper articles from both Dutch and Belgian newspapers from 1993-2005. The spoken genre takes its words from the CGN (Corpus Gesproken Nederlands), which is made up of 900 hours of spoken Standard Dutch from 1998-2004 (ca. 9 million words). The web corpus also uses SoNaR, but concentrates on blogs, discussions, e-magazines, newsletters, press releases, and Wikipedia pages (Tiberius & Schoonheim, 2014).28 In the event that a word appears across more than one genre, the location of higher frequency determines its placement in the dictionary's four genres. For example, should

28 For this test, both the spoken and web genres did not contribute any results.
a word appear in the newspaper corpus twice as much as in another corpus, it is placed in the newspaper genre (Tiberius & Schoonheim, 2014). For this first study of test two, 12 verbs for the split form and 9 for the unsplit were not found in the frequency dictionary. Despite this slight lack, the Frequency Dictionary furnishes useful results.

Instead of observing only the "met_verbs", test two observes all verbs in the data set. Modal verbs and auxiliaries are still removed from the data, for reasons stated above in test one. Each verb is marked for being inside or outside of the split or unsplit construction, following the methodology previously set forth. A new level of complexity enters in test two, whereby one must also take account the verbs' genre/list placement and how this affects the analysis. As such, this test observes the frequency of all used verbs with their respective $p_{\log 10}$ values, and considers how genre/list categorization could affect the understanding for the motivation in a usage variation between the split or unsplit forms. Thus, test two builds upon the methodology from test one and suggests a possible motivation for this difference, namely, the frequency of the verb's use as well as the genre to which it belongs.

6.3 Genres/Lists

For the split forms, there is an equal number of verbs with positive $p_{\log 10}$ values belonging to both the general and core vocabulary lists.
Figure 3: All verbs in the split form with positive $\log_{10}$ values

Since the Core vocabulary list represents the most common words in the language, it is interesting to note how many of these items are attracted to the split form. With the general list, the situation is similar. These are words that appear across three or more genres, but are not part of the core vocabulary. They also make up an equally large portion of the verbs in the split form, yet their frequency values out of 100 texts is not be as high as those in the core list. Verbs belonging to the Fiction and Newspapers genres make up a smaller portion of those verbs attracted to the split form. The average frequencies and positive $\log_{10}$ values for the split form are below.
<table>
<thead>
<tr>
<th>Genre/List</th>
<th>Frequency</th>
<th>$p_{log10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>32.17</td>
<td>3.40</td>
</tr>
<tr>
<td>Fiction</td>
<td>2.36</td>
<td>3.32</td>
</tr>
<tr>
<td>General</td>
<td>3.84</td>
<td>1.28</td>
</tr>
<tr>
<td>Newspapers</td>
<td>6.59</td>
<td>2.06</td>
</tr>
</tbody>
</table>

Table 11

For the split values the Newspaper genre, whereas only 13% of the total positive $p_{log10}$ values, maintains the second highest frequency per 100 documents. Then again, the Newspaper genre also has the second highest frequency value of the lists/genres all together at 66.8 appearances out of 100 documents. These data would suggest that more frequently occurring verbs may actually be attracted to the split form.

For the unsplit form, the results of the lists/genres distribution among the positive $p_{log10}$ values differs significantly. Here, the General list dominating with over half of the numbers, and the core vocabulary is merely a quarter of the total as opposed to close to half.
The Newspaper and Fiction genres remain minimally changed. The table below shows the average frequencies and $p_{\log10}$ values of the verbs with positive values in the unsplit and split forms. We see that the split form has higher $p_{\log10}$ values across all four lists/genres and also higher frequencies in the core and newspapers categories. Since these two categories have the first and second highest frequency value, respectively, it could further be suggested that more frequently occurring words will tend to appear with the split form. In the unsplit forms, the general and fiction categories have slightly higher frequency averages, but they still maintain considerably lower $p_{\log10}$ values.
<table>
<thead>
<tr>
<th>Genre</th>
<th>Frequency</th>
<th>$p_{\log_{10}}$</th>
<th>Frequency</th>
<th>$p_{\log_{10}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>32.17</td>
<td>3.40</td>
<td>27.82</td>
<td>1.06</td>
</tr>
<tr>
<td>Fiction</td>
<td>2.36</td>
<td>3.32</td>
<td>3.32</td>
<td>0.61</td>
</tr>
<tr>
<td>General</td>
<td>3.84</td>
<td>1.28</td>
<td>4.27</td>
<td>1.09</td>
</tr>
<tr>
<td>Newspapers</td>
<td>6.59</td>
<td>2.06</td>
<td>4.45</td>
<td>1.57</td>
</tr>
</tbody>
</table>

Table 12

Therefore, one could state, based on the data presented above, that more frequently occurring verbs should appear with the split form, and less frequent with the unsplit. Tying into the earlier statements regarding idiomatic expressions in test one, I suggest that more frequent verbs may show a likelihood to occur more with the split form, and less frequent with the unsplit one. Since words with a higher frequency are more easily parsed in the mind ahead of time without need to hear the whole expression (i.e. the word frequency effect), a similar situation may be happening with the split form. It isn't necessary to keep the pronominal adverb together as one unit, as in the unsplit form, as the listener/reader will be quicker to parse the "missing" mee that appears at the end of the clause. This phenomenon is also discussed in (DeLong, Urbach, & Kutas, 2005), mentioned above. In the following part of test two, I test for correlation between frequency, $p_{\log_{10}}$ values, and list/genre category.

6.4 Correlation

This section of test two uses Pearson's R to determine if there is a correlation between the individual verb's frequency and attraction or repulsion to each construction. To begin, each verb needed to be tested for its $p_{\log_{10}}$ value. This process was simply continued from test one, but extended to all verbs inside and outside of the two constructions. Once completed, the next step
was to graph out the verbs frequency with respect to their $p_{\log 10}$ values. Using scatter plots where the x axis represents the $p_{\log 10}$ values (attraction/repulsion to construction) and the y represents the frequency per 100 documents, one can see that there is basically no correlation between the split form and its relation to frequency (Figure 5) but there is a slight negative correspondence between increased frequency and $p_{\log 10}$ values for the unsplit form (Figure 6). This shows us that the higher the frequency, the less likely there appears any attraction to the unsplit form.
Figure 5: $p_{\log 10}$ values (x) vs. frequency (y) in split construction
Figure 6: p_{log10} values (x) vs. frequency (y) in unsplit construction
Pearson's R provides a value of $R = -0.024031524$ for the split forms; a very low correlation indeed. The trend line, visible as the series of dots labeled as "Linear (Split)", also shows a near 0 correspondence. In other words, the $\text{plog10}$ values and the frequency appear to have no effect on the split form whatsoever. For the unsplit form, however, the situation is different. Whereby the split form shows nearly no correlation at all, the unsplit form delivers a negative result from Pearson's R of $-0.237150113$. This negative correlation speaks to the hypothesis that frequency affects the attraction/repulsion to the constructions. Therefore, I modify hypothesis$_1$ to state:

$$H_0: \text{There is no factor contributing to the difference between split and unsplit forms.}$$

$$H_1: \text{The higher a verb's frequency, the less likely it will appear with the unsplit form.}$$

Yet such a hypothesis is somewhat problematic, as grammars of Dutch claim the unsplit form to be the "standard" (Bilderdiijk, 1826).\footnote{Bilderdiijk's original text reflects the opposite. Splitting should be considered the standard, in his opinion: De ware Hollander zegt, WAAR ik zeer VAN verwonderd was, WAAR ik my hartelijk IN verheugde; WAAR ik OM gehaat werd; en zoo ook: HIER neem ik deel IN, HIER maak ik een besluit of gevolg UIT op; DAAR ga ik MEE naar bed, DAAR moet gy genoegen MEE nemen: en dwaaslijk willen de schoolvossen van de later tijden: WAARVAN ik verwonderd was, WAAROVER ik verblijd was, WAARIN ik my verheugde, WAAROM ik gehaat werd [...]. Lafa gerammel van die noch gehoor, noch gevoel, noch verstand hebben voor de samenvlechting eener rede! Niet, dat het waaraan, hierin, wherethrough, in zich-zelfs kwaad is, maar zich daaraan te binden is stijve eenvormigheid [...].}

If the unsplit form were to be affected by the verb's frequency, and the split remains unaffected, one could then assume that the split form is the "standard" whereas the unsplit is used (when the form can be split) to signal something.

The true Dutchman says: WHAT I was very amazed ABOUT, WHAT I am heartily pleased ABOUT; WHAT I am hated FOR; and furthermore: HERE is something I'm taking part IN, [HERE] I make a decision or result OUT of it; I'll go to bed WITH it, it's something you have to get enjoyment out OF; and unwisely the school teachers of the latter days proclaim: WHEREABOUT I was very amazed, WHEREOVER I was very happy, WHATABOUT I am pleased, WHEREFOR I am hated [...]. Utter nonsense from those who have neither the ear nor the feeling, nor the understanding of the interweaving of a speech! Not to say that whereupon, herein, wherethrough are wrong themselves, but to hold oneself to such forms is stiff uniformity. [translation: de Carlo] (Bilderdiijk, 1826, p. 349).
This negative correlation between higher frequency and lower $p_{\text{log10}}$ values for the unsplit form is also visible when analyzing each of the four lists/genres. Only the General list provides a slightly positive correlation, and fiction displays an extremely small positive correlation.

<table>
<thead>
<tr>
<th></th>
<th>Split</th>
<th>Unsplit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>-0.02</td>
<td>-0.24</td>
</tr>
<tr>
<td>Core</td>
<td>0.03</td>
<td>-0.34</td>
</tr>
<tr>
<td>General</td>
<td>-0.11</td>
<td>0.17</td>
</tr>
<tr>
<td>Fiction</td>
<td>-0.16</td>
<td>0.01</td>
</tr>
<tr>
<td>Newspaper</td>
<td>-0.25</td>
<td>-0.30</td>
</tr>
</tbody>
</table>

Table 13

The unsplit forms, overall, show a stronger negative correlation between higher frequency and attraction. The variation with the Fiction genre is so minimal that one could argue that there is no correlation at all. One issue when analyzing the Fiction and Newspaper genres pertains to the amount of positive $p_{\text{log10}}$ value data points. For Fiction, there are have 10 for the split form, 38 for the unsplit. In the Newspaper genre, 14 appear for the split and 62 for unsplit. These values stand in stark contrast to the Core and General lists: 52 in General for split, 204 unsplit and Core exhibits 128 for split and 182 for unsplit. It would be interesting to observe how a much larger data set with more verbs belonging to the Fiction and Newspaper genres would react. Yet for now, it appears that the split form continues to reject more difficult vocabulary (hence the negative correlation with Fiction) and attracts more basic vocabulary (positive correlation with the Core list).
6.5 Conclusion Test Two

In concluding this test, it can be argued on the one hand that there is a slight positive correlation between more frequent verbs (Core list) and increased p\text{log}_{10} values in the split form. On the other hand, frequently occurring verbs tend to be less attracted to the unsplit form. These correlations can then be further understood along the lines of phrasal and verbal idioms. If phrasal/verbal idioms are more easily processed than non-phrasal/verbal idioms, and they have a tendency to appear with more frequently occurring verbs, then one would expect these frequently occurring verbs to contain less semantic information. They might not be used to signal a significant event in the discourse, or they might not direct the attention of the reader/listener to something. Since I see these appearing with the split form, I could further speculate that the split form signals little to nothing semantically. The unsplit form, by contrast, repulses verbs of high frequency. The listener/readers is not able to simply gloss over the information or parse it ahead of time, as may be the case with the split form.

Consequently, the unsplit form, with its infrequent verbs, may signal something indeed. Via metaphorical extension, one can understand the two forms as traffic signals. The split form functions as a yellow light. The driver may need to slow down, yet depending on his/her distance from the stop light, he/she can continue along in the same direction, unhindered. The unsplit form, by contrast, is more similar to a red light. The driver needs to come to a complete stop, and the reader/listener is confronted with a larger chunk of information at the onset and must stop and observe his/her surroundings. Only when the light turns green, or rather when the information continues with the unsplit form, can the listener/reader continue and determine which direction to go. To tie this back into Bolinger's Linear Modification (1952), the split form does not narrow down the possibilities of discourse, it keeps them open, perhaps to continue
what was previously spoken. Use of the unsplit more instantly narrows the discourse possibilities, and centering on what may come ahead. The reader is directed to focus on the following topic.

The results from test two only hint at a possible way of finding a meaning difference between the two forms. We have seen how frequency can have an effect on choice between the two forms and have yet to postulate a true, quantifiable meaning. In the following test three, I use Princeton's WordNet database to observe just how much semantic information appears with the verbs in the split and unsplit forms. The results from test three shall help form a semantic interpretation between the two forms.
7 Test Three: WordNet & Transitivity

Test two demonstrated how frequency can have an effect on the choice of one form over the other. What remains is to determine the semantic manifestation of this choice of less frequently occurring verbs appearing alongside the unsplit. For if one determines a reason for choice, based solely on quantitative values, one still overlooks the one aspect of language that makes it innately human: meaning. What does this difference in choice signal when discussing meaning difference? Can a meaning difference be determined? Test three tackles these questions and suggests a difference based on yet more quantitative data.

Since test one showed that the choice in split or unsplit is not dependent on the verb's attraction to the pronominal adverb, but rather on the frequency of the verb (test two), the next step is to determine what kind of verbs these are. Utilizing Princeton's WordNet database, test three categorizes each individual verb according to a semantic category provided by WordNet. Test three also questions what trends appear when taking these semantic categories into account. Do these trends speak to a semantic difference between the two forms, and if so, is it significant enough to justify positing a difference? In order to provide further quantitative grounding for the following hypothesis, I utilize the Hopper and Thompson's Transitivity hypothesis (1980). This will be defined and discussed further below.

The verb's semantic category provided by WordNet serves as the quantitative test which will inform the following hypotheses:

\[
H_0: \quad \text{No difference in semantic category between the two forms can be determined}
\]

\[
H: \quad \text{Verbs occurring with the unsplit form will show a stronger likelihood to use semantic categories that contain verbs of lower transitivity, whereas verbs in the}
\]
split form will exhibit a tendency to co-occur with verbs that are more likely to be higher in transitivity.

7.1 Methodology

For the third quantitative test, I make extensive use of Princeton's WordNet database. WordNet was conceived in 1985 at Princeton's Cognitive Science Laboratory. Initially, WordNet was to "identify the most important lexical nodes by character strings and to explore the patterns of semantic relations among them" (Fellbaum, 1998, p. xvii). Since the initial goal of WordNet was to create semantic relations, definitions were considered superfluous; one could infer the meaning by observing the network. These relations are organized in synonym sets (synsets) of nouns, verbs, adjectives, and adverbs. Each member of the synset is related to the main concept, but cannot be necessarily be interchanged with other members (Fellbaum, 1998). In cases where the word can be seen as polysemous, it is included in multiple synsets, as in {trunk, tree trunk}, {trunk, torso}, and {trunk, proboscis} (Fellbaum, 1998). Yet, not only are the individual part of speech categories related, but also their counterparts in other fields, such as interrogation, interrogator, interrogate, interrogative (Fellbaum, 1998).

Verbs are organized foremost troponymically, that is where one verb describes how another is done, i.e. {whisper}-{talk} or {punch}-{strike}. A secondary method of organizing verbs is via backward entitlement (divorce-marry), presuppositions (buy-pay), and cause (show-see) (Fellbaum, 1998). So, for each verb extracted in the data, not only did WordNet provide a primary meaning, but also other related words. This was extremely helpful, as WordNet is in English and the data are in Dutch. Currently, there are 13767 synsets and 25947 word-sense

---

30 Over time, this changed: definitions were to be included in later iterations of WordNet.
31 More on the idiosyncrasies of this issue are discussed below.
pairs in the verb category, of these 6277 are monosemous words/senses and 5252 polysemous words and 18770 polysemous senses.

Beyond the internal part of speech organization, each verb in WordNet can belong to one of 15 different semantic categories\textsuperscript{32}:

<table>
<thead>
<tr>
<th>Semantic Category</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>verb.body</td>
<td>verbs of grooming, dressing and bodily care</td>
</tr>
<tr>
<td>verb.change</td>
<td>verbs of size, temperature change, intensifying, etc.</td>
</tr>
<tr>
<td>verb.cognition</td>
<td>verbs of thinking, judging, analyzing, doubting</td>
</tr>
<tr>
<td>verb.communication</td>
<td>verbs of telling, asking, ordering, singing</td>
</tr>
<tr>
<td>verb.competition</td>
<td>verbs of fighting, athletic activities</td>
</tr>
<tr>
<td>verb.consumption</td>
<td>verbs of eating and drinking</td>
</tr>
<tr>
<td>verb.contact</td>
<td>verbs of touching, hitting, tying, digging</td>
</tr>
<tr>
<td>verb.creation</td>
<td>verbs of sewing, baking, painting, performing</td>
</tr>
<tr>
<td>verb.emotion</td>
<td>verbs of feeling</td>
</tr>
<tr>
<td>verb.motion</td>
<td>verbs of walking, flying, swimming</td>
</tr>
<tr>
<td>verb.perception</td>
<td>verbs of seeing, hearing, feeling</td>
</tr>
<tr>
<td>verb.possession</td>
<td>verbs of buying, selling, owning</td>
</tr>
<tr>
<td>verb.social</td>
<td>verbs of political and social activities and events</td>
</tr>
<tr>
<td>verb.stative</td>
<td>verbs of being, having, spatial relations</td>
</tr>
<tr>
<td>verb.weather</td>
<td>verbs of raining, snowing, thawing, thundering</td>
</tr>
</tbody>
</table>

Table 14

For each item in the database, WordNet assigns a semantic category. It is possible that a verb can be polysemous, as mentioned above. In this case, the verb is tagged with the semantic category to which it may belong, depending on context. These semantic categories were initially assigned via intuition in WordNet, yet over time this method has fallen out of favor, as there are currently many more example sentences available through internet data (Fellbaum, 1998) This test uses these 15 semantic categories as the foundation to determine if there may be a meaning difference between the two forms.

\textsuperscript{32} There are 45 total semantic categories for all four parts of speech. Since I am only observing verbs here, these will be omitted.
7.2 Data Preparation

The first step in test three was to extract a random sample of the data, since analyzing each verb for its semantic category in the mass data set was too time consuming. The data were randomized and 500 examples were initially chosen. We chose a sample size of around 440 clauses for each construction. The number was narrowed down to 440 after removing sources that appeared multiple times, as this may have provided issues with a random distribution of the data. From these 440 examples, the auxiliaries *zijn, worden* or modals were extracted, leaving with 281 examples in the split form and 336 in the unsplit.33

After this, I moved to proper cataloguing of all verbs inside and outside of the construction. Not only infinitives were counted, but all verbs that occurred as phrasal verbs or in idiomatic expressions. With the verbs inside and outside of the constructions extracted, the following steps included considerable amounts of translation. Since WordNet is in English, each verb had to be translated first into Dutch.34 These translations were completed via the van Dale online Dutch-English dictionary (van Dale, 2017). Seeing that a majority of the verbs could have multiple senses, especially in translation, it was imperative to test each one with regard to the context provided in the corpus extract. This added time to test three, yet was important to ensure proper labelling. These senses also belong to one of the aforementioned 15 semantic categories. Should the verb belong to more than one semantic category, the best fit was chosen. From here, the data were organized according to their category and counted. No advanced statistical tests were run on test three, only percentages. This decision was made as the main goal of test three is

33 Since the data in this set were smaller than previous tests, instances of *hebben* remained, but instances where *hebben* functioned as an auxiliary were removed.

34 At the time of the completion of test three, there was no readily available open source Dutch lexical word net. As of June 4, 2016, there is currently an open source Dutch word net available here: http://wordpress.let.vupr.nl/odwn/
to see if WordNet's semantic categories could provide any insight into the discrepancies found in both tests one and two between the two forms.

7.3 Results

For the split form, the data show a strong preponderance of stative verbs (28%). Recall that all instances of *zijn* and *worden* have been removed from the data. This alone is interesting, as stative verbs, by their very nature, do not express dynamism— they are static. The graph below illustrates the percentages of semantic categories in the 281 instances used for the split construction.

![Figure 7: Percentage of semantic categories in split construction](image)

Figure 7: Percentage of semantic categories in split construction
Two variations of the verb phrase *mee te maken* were the foremost in the list of stative verbs. One variation with 23 occurrences is *mee te maken hebben* and the second variation is *mee te maken krijgen* at 7 occurrences. Both verbs do not particularly highlight any immediate action, as they support information that has previously appeared. For example, the phrases with the split pronominal adverb below function anaphorically.

(25) Het ging juist om de toegang tot de Marokkaanse wateren voor de Spaanse en Portugese vissers. **Daar heeft Nederland niets te maken.**

'It concerns precisely access to Moroccan waters for Spanish and Portuguese fishermen. The Netherlands has nothing to do with that.'

(26) Plomp zegt zich nog niet georiënteerd te hebben op de grens-problematiek tussen Meppel en Staphorst, maar hij zal **daar zeker nog veel mee te maken krijgen.**

'Plomp states that he had not yet addressed the border issue between Meppel and Staphorst, but he will most likely have a lot more to deal with regarding that.'

Example (25) points back at the information just discussed, namely that the point of discussion was how Spanish and Portuguese fishermen were to handle access to Moroccan waters, and this point of contention does not affect the Netherlands. The information contained in the split construction is not new, it only underlines the static fact that the Netherlands remains untouched by this ordeal. Example (26) looks forward with the future auxiliary *zullen*, yet maintains an anaphoric and static reference. Since *krijgen* translates as "to get" or "to receive" in English, there are two options available for the WordNet analysis. Often, the verb has the semantic category of possession, yet this is a case where context was extremely important. Of the two
translations, "to receive" functioned best, and in this sense, the context expressed experiencing something as a reaction to what happens. We can see this in the example above. Plomp does not show himself to be focused on the border problem between Meppel and Staphorst, but he will have to become experienced in dealing with this issue. Therefore, *krijgen* in this verb construction appeared in the static semantic category for verbs, as each act of receiving was a reaction to happened in the previous clause. For each instance of *mee te maken krijgen* I tested this semantic category, and it functioned best as a stative verb.

The second highest semantic category is cognition. Similar to stative verbs, cognitive verbs do not express a particularly high level of dynamism. In this category, the verb phrases *rekening houden* 'to take into account' and *moeite met iets hebben* 'to have difficulty with something' appear most frequently, at 27 and 25 occurrences, respectively. Example (27) below exhibits how *rekening houden* often appears throughout the data.

(27) Het sexuele misbruik zou in 1983 gestopt zijn toen de man trouwde. Zelf is D. ook misbruikt door zijn oudere broer. **Ds rechtbank hield daar rekening mee.**

'Supposedly the sexual assault stopped in 1983 when the man was married. Even D. was abused by his older brother. D's court kept that in mind.'

As is the case with the previous semantic category, stative, the split constructions here function anaphorically, not providing the discourse with any particularly dynamic information. To take something into account requires that the *something* taken into account be either known or relevant to the discussion at hand. In the example above, the writer describes how the court in D's case took into account that D. was also sexually abused by his elder brother. The court took
this fact into account, referring anaphorically to the abuse, and therefore not including any new information. The clause is flat and does not move the topic forward.

The second most frequent verb expression, *moeite met iets hebben*, also does not provide interesting material by way of semantic weight. In example (28) below, half of the viewers do not have any problem finding the desired program, a mental activity.

(28) Het maakt de kijkers ook niet uit wie er uitzendt, zolang hij het gewenste programma maar kan vinden. Ongeveer de helft heeft **daar geen moeite mee**.

'The viewers do not care what is broadcast, as long as they can find the desired program. Around the half of them have no trouble at all in doing so.'

Among the other examples of *moeite met iets hebben*, no foregrounding information was referenced. 'To have', *hebben*, does not require much imagination on the part of the listener/reader and the action involved is easily comprehended. One either has something or does not, and whereas *hebben* on its own is placed in the semantic category of possession, when used in conjunction with *moeite* it shifts to the realm of cognitive verbs. For when one has difficulty with something non-physical, the object of difficulty becomes an intangible challenge (particularly in the examples in the data), and must be overcome through mental processes.

Stative and cognitive verbs account for over one half of all of the verbs found in the constructions for the split data set. These two semantic categories do not tend to involve multiple participants, and in many cases only the subject. Expressing either a static moment (stative verbs) or moments of internal reflection (cognitive), the action appears flat. The reader/listener does not confront an action requiring much attention.
This stands in stark contrast to the predominant categories found in the unsplit form. In the figure below, it is clear that verbs of change and communication make up over a quarter of the unsplit form's verbs.

![Unsplit: WordNet Percentages](image)

**Figure 8: Percentage of semantic categories in unsplit construction**

The two most frequent verbs in the change category are *brengen*, 'to bring', and *komen*, 'to come'. Recall that the infinitives themselves did not determine the semantic category, but each individual verb's use in context. Therefore, if a verb had multiple semantic categories, each one was considered before assigning it to the verb in that sentence. For example, in instances of *komen* the context expressed a clear change. The verb often did not function merely as one showing physical movement toward an object, but rather change, as seen below.

(29)  De sensatie van de avond was de overwinning van Hans Renes. Daarmee **kwam** de stand weer in evenwicht.
'The highlight of the evening was Hans Renes's victory. With that, equilibrium returned.'

In the example above, the sentence before the unsplit construction sets the scene for something drastic. There was a sensation that evening, for Hans Renes triumphed. What then transpired after this victory was the restoration of balance, there was a change. The *stand*, i.e. the score, did not physically move, but rather changed. With *komen*, the verb appeared throughout the data functioning not as a verb of motion but rather one expressing change.

1. *hoger te liggen komen* (to come to lie higher than before)
2. *in aanmerking komen* (to qualify for something)
3. *onder IFOR vlag komen* (to come under the flag of IFOR/NATO)
4. *tot stand komen* (to materialize)
5. *op [monetary sum, ranking] komen* (to raise to a level)
6. *een eind aan iets komen* (to come to the end of something)

The other major verb, *bren
gen*, co-occurs with expressions such as:

1. *een nieuw element in de discusie brengen* (to bring a new element into the discussion)
2. *op [point, sum] brengen* (to bring to a level)
3. *gelijk brengen* (to bring to equilibrium)

From the list above, it is clear that the context was key in understanding the uses of *bren
gen* and *komen*. Alone, it belongs to the semantic category of motion, yet in conjunction with the phrases, the simple verb takes on entirely new senses. Hence, *komen* and *bren
gen* are no longer used in their literal sense, but more figuratively. Figurative use of verbs generally requires more mental parsing, and therefore creates a more vivid image in the reader's/listener's mind.
Other verbs found in the unsplit construction also point toward more figurative use. *Doelen* (to aim) does not appear in the literal sense of aiming a weapon at a target, but rather aiming at a topic, to communicate with another party the speaker's/writer's goal. *Staan* (to stand) does not take on the literal sense of physically standing, but rather standing *op het spel* (to jeopardize, to menace) or *op de tocht* (to cancel). In both cases *staan* is not standing by, statically and passively, but rather signaling a change, telling the reader/listener that the activity is reaching a turning point. *Doelen* does not aim in competition, but rather communicatively, driving the discourse forward at an object in the distance. We then see that verbs occurring inside the unsplit construction have a higher tendency to belong to semantic categories of change or communication—two categories that foreground more than background, as they often provide new information to the listener/reader.

### 7.4 Transitivity

Semantic categories on their own can say much, but in this case, it is important to find another method to quantify the results further. Hypothesis one states that verbs occurring with the unsplit form will show a stronger likelihood to use semantic categories that contain verbs of lower transitivity, whereas verbs in the split form will exhibit a tendency to co-occur with verbs that are more likely to be higher in transitivity. To understand the Transitivity Hypothesis is to understand how the frequency of each of these semantic categories bears on the meaning difference discussed in the following section.

The Transitivity Hypothesis grounds itself on the notion that verbs can be analyzed on 10 various properties. Each of these properties displays how the "effectiveness or intensity with which the action is transferred from one participant to another" (Hopper & Thompson, 1980, p. 252). Hence, Transitivity shows how much the action of the verb affects its objects; the higher
the amount of transitive properties a verb has, the higher the correlation with foregrounding. Conversely, the lower the transitivity, the higher the correlation with backgrounding (Hopper & Thompson, 1980). The 10 components and their parameters values are seen below, as shown by Hopper and Thompson (1980).

<table>
<thead>
<tr>
<th>Component</th>
<th>High Transitivity</th>
<th>Low Transitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>2 or more participants, Agent and Object</td>
<td>1 participant</td>
</tr>
<tr>
<td>Kinesis</td>
<td>action</td>
<td>non-action</td>
</tr>
<tr>
<td>Aspect</td>
<td>telic</td>
<td>atelic</td>
</tr>
<tr>
<td>Punctuality</td>
<td>punctual</td>
<td>non-punctual</td>
</tr>
<tr>
<td>Volitionality</td>
<td>volitional</td>
<td>non-volitional</td>
</tr>
<tr>
<td>Affirmation</td>
<td>affirmative</td>
<td>negative</td>
</tr>
<tr>
<td>Mode</td>
<td>realis</td>
<td>irrealis</td>
</tr>
<tr>
<td>Agency</td>
<td>agent high in potency</td>
<td>agent low in potency</td>
</tr>
<tr>
<td>Affectedness of Object</td>
<td>object totally affected</td>
<td>object not affected</td>
</tr>
<tr>
<td>Individuation of Object</td>
<td>object highly individuated</td>
<td>object non-individuated</td>
</tr>
</tbody>
</table>

Table 15

For each of these components, their parameters are discussed at length in Hopper and Thompson (1980) and I would direct the reader to consult there for an in-depth discussion of each component and parameter.

After analyzing the results from the WordNet test, I assign high or low transitivity values for each semantic category. This was done by extracting, when possible, three random verbs from each semantic category which occur in the split or unsplit form. Although verbs were chosen at random, it was important to first look at the corpus results in further detail. When
delivering the results from the 38-million-word corpus, I set the KWIC (key word in context) limit to seven words on either side, and in order to test the transitivity properties for each verb, it was necessary to ensure a full clause. Verbs that did not provide enough information (subject, object, time adverbials) were omitted. For this reason, some semantic categories only provided two verbs.

Verbs in the split and unsplit form around 30 each. With each verb, I extract the clause in which they appear and test the verbs for high or low transitivity based on the parameters above. If the verb in the clause, for example, has two or more participants, I mark as having a positive value. Should the verb only require one participant, then I mark it as having a negative value. This is followed throughout all of the parameters in Table 15. In the event that it was unclear whether one parameter is higher than the other, I marked these as well. Therefore, for this test, each verb is marked for high transitivity, low transitivity, or if it was unclear transitivity. The unclear values are left out of this discussion. Should a particular semantic category exhibit higher transitivity, I shall adopt Hopper and Thompson's statement that higher transitivity correlates with foregrounding and lower transitivity with backgrounding.

First, the split forms display a strong predilection for verbs belonging to the stative class, over one quarter- 28%. These verbs are also lowest in transitivity, as 77% of all transitivity properties were negative. Hence, stative verbs appear inherently low in transitivity, suggesting more background information and less required cognitive effort. Cognition, the second highest semantic category in the split forms at 25%, is also the second highest in negative properties at 67%. These two semantic categories, stative and cognition, make up 63% of all verbs in the split form for this test and rank highest in low transitivity. Therefore, I state that split forms, with

---

35 The average percentage of all unclear values in both split and unsplit forms is 1.4%.
36 Cognitive verbs ranked 100% negative in the parameters of Kinesis, Aspect, Affectedness of O, and Individuation of O. For Volitionality, Punctuality, and Mode, two thirds of the values were negative.
their respective semantic categories, display lower transitivity, supporting a suggestion that one should not expect any change in discourse, any new information to be brought about in split forms.

Figure 9: Pct. of pos. & neg. values for transitivity per semantic category: split

By contrast, the unsplit forms show how its highest ranking semantic category, change, displays astoundingly high positive values for transitivity. Seeing that verbs of change mark an event that should be recognized by the reader/listener as noteworthy, there is an expectation of high transitivity values, since, according to Hopper and Thompson (1980), high transitivity = foregrounding. This test delivers such a result. At 25% of all verbs in the unsplit form, the semantic category of change comprises a significant amount. For the verbs tested here, 83% of

---

37 The split forms only have 4.3% of all verbs belonging to the change category, by contrast.
the transitivity properties are positive, leading to further evidence for the thesis behind this dissertation: There is no synonymy of meaning here.

Figure 10: Pct. of pos. & neg. values for transitivity per semantic category: unsplit

The following section will suggest a difference in meaning between the two forms, based on the data provided here. Via transitivity, the difference in how the split and unsplit forms choose their verbs. Since the split forms make use of more verbs of semantic categories that have lower transitivity, one can see how backgrounding would be one factor for splitting. The unsplit form, however, shows a strong tendency to use verbs belonging to the change category. These verbs are also much higher in transitivity in the unsplit form. This is also seen in the total values for each semantic category across both split and unsplit forms in Figure 11, below. Stative verbs contain a much lower number of positive transitivity properties, and verbs of change more
positive. Therefore, the split form will display a stronger tendency to background information and the unsplit to foreground new information.

Figure 11: Pct. of pos. & neg. values for transitivity per semantic category: split & unsplit

These values are presented below in tabular form. The change category is ranked much lower in the split form than in the unsplit.
<table>
<thead>
<tr>
<th>Rank</th>
<th>Split Category</th>
<th>Rank</th>
<th>Unsplit Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stative</td>
<td>1</td>
<td>Change</td>
</tr>
<tr>
<td>2</td>
<td>Cognition</td>
<td>2</td>
<td>Communication</td>
</tr>
<tr>
<td>3</td>
<td>Possession</td>
<td>3</td>
<td>Stative</td>
</tr>
<tr>
<td>4</td>
<td>Communication</td>
<td>4</td>
<td>Possession</td>
</tr>
<tr>
<td>5</td>
<td>Social</td>
<td>5</td>
<td>Social</td>
</tr>
<tr>
<td>6</td>
<td>Competition</td>
<td>6</td>
<td>Motion</td>
</tr>
<tr>
<td>7</td>
<td>Change</td>
<td>7</td>
<td>Perception</td>
</tr>
<tr>
<td>8</td>
<td>Motion</td>
<td>8</td>
<td>Cognition</td>
</tr>
<tr>
<td>9</td>
<td>Creation</td>
<td>9</td>
<td>Competition</td>
</tr>
<tr>
<td>10</td>
<td>Contact</td>
<td>10</td>
<td>Contact</td>
</tr>
<tr>
<td>11</td>
<td>Perception</td>
<td>11</td>
<td>Creation</td>
</tr>
<tr>
<td>12</td>
<td>Body</td>
<td>12</td>
<td>Emotion</td>
</tr>
<tr>
<td>13</td>
<td>Consumption</td>
<td>13</td>
<td>Body</td>
</tr>
<tr>
<td>14</td>
<td>Emotion</td>
<td>14</td>
<td>Consumption</td>
</tr>
<tr>
<td>15</td>
<td>Weather</td>
<td>15</td>
<td>Weather</td>
</tr>
</tbody>
</table>

Table 16: WordNet Semantic Category Ranking (Based on Percentage of Total)

7.5 Meaning Difference

It is this crucial aspect of the analysis, determining context, that requires a trained eye. Computational analysis can bring the linguist so far, but it is not until the step is taken to dive deeper into the context that patterns in meaning may emerge. For this reason, test three does not utilize advanced computational methods to suggest meaning, as meaning is an innate human characteristic. Counts and percentages provide an overview of the data, yet it is the cautious, introspective linguist who must accept these overviews with caution. Test one provides strong evidence against van der Horst's (1992) theory. Test two underlines a correlation between less frequently used words and the unsplit form. To have less frequently occurring verbs in a construction should signal something to the listener/reader, that some action that may be considered more salient, for example. As test three showed, the unsplit form also makes dominant use of verbs expressing change and communication, whereas the split form uses verbs
expressing states and cognition. In addition to the semantic categories of the verbs in test three, the predominant semantic categories in each form display either high transitivity (unsplit forms) or low transitivity (split forms). It would then appear that the two constructions embody polar opposites of each other. Where the unsplit changes, the split remains the same.

A difference in meaning has hitherto not been suggested for the choice between the two forms, yet I shall suggest that there is indeed a difference in meaning. Test two has laid the quantitative foundation for a reason behind the choice between the two forms. There is a stronger tendency for less frequent verbs to collocate with the unsplit construction. Yet this alone is not enough to express a difference in meaning; it is just a motivating factor. When speakers have two options on hand to express what appears to be the same, they can often make a subconscious choice for one over the other. This preference may not be entirely clear, and a meaning difference may seem too slight to notice, nevertheless a difference is there.

We suggest that the split form does not signal any change in discourse, but rather maintains continuity. The split form does not foreground any new information, but rather remains static or provides background. Low transitivity values for the split form also support this backgrounding hypothesis, for if the verbs were to be higher in transitivity, I would expect more foregrounding (Hopper & Thompson, 1980).

In contrast to the split form, the unsplit signals a possible shift in discourse, a form of discontinuity. By taking the higher levels of transitivity into account among the highest ranking semantic category (change) of the unsplit form, one can also observe how more foregrounding would be expected to occur in unsplit forms. The reader/listener is also cued into this possible change via the manifestation of the unsplit pronominal adverb. The word itself is longer than the

---

38 The one difference is stated above, namely that of Donaldson (2008). His suggestion, however, is the opposite of what I am suggesting here. In fact, I argue that his meaning difference is better suited for German, as German also has this construction, yet it is considered non-standard and is left out of further discussion for this reason.
split and functions to direct the attention forward, similar to the concept of linear modification, mentioned in test two (Bolinger, 1952). Immediately, the reader/listener is honed on the topic with the unsplit form, the information is narrowed down and he/she is signaled that something more salient will be coming sooner than later. We see this with the unsplit form quite clearly. Taking example (29) from earlier, one can imagine how the focus is narrowed:

(29) Daarmee kwam de stand weer in evenwicht

Figure 12: Bolinger's linear modification

Via Bolinger's figure above as a reference, one can visualize how the important information is focused at the end. With each step, the reader comes closer to understanding what happened. It is not until the sentence is complete that the reader can parse the expression *in evenwicht komen* and fully understand how the unsplit form is grabbing the reader's attention to focus on what will happen ahead.

To revisit the traffic analogy suggested in chapter 6, the split form works semantically like a speed bump on a neighborhood street. The driver continues in the same direction, but he may be slowed down a little bit. With the unsplit form, the driver is at a four way stop sign. He must first come to a complete stop and has the option to go in three directions – to the right, the left, or continuing in the same direction. This happens only after stopping to look what may be coming from either side. Thus, I argue that the split form exhibits a use congruent with continuous discourse and stasis, whereas the unsplit displays a use embodying shift in discourse and dynamism. In this way, I question the prevailing theories that believe there to be no difference at all between the two forms, and endeavor to provide a new reading of the two forms.

In the final test, I will utilize two novels from Netherlandic Dutch author, Willem Frederik
Hermans, in an effort to underline how the two forms differ in meaning. The extracts will show how foregrounding and backgrounding work within each passage. Yishai Tobin's concept of Semantic Integrality further supports $H_1$ suggested in test three.
8 Test Four: Literature & Semantic Integrality

The final test of this dissertation involves using the hypotheses laid out in the last two tests, Yishai Tobin's concept of Semantic Integrality (Tobin, Semiotics and Linguistics, 1990) and two works of literature from a renown Dutch author, Willem Frederik Hermans. Hermans is well known as one of the foremost Netherlandic Dutch authors of the previous century, and the two works chosen here, *Nooit meer slapen* (1966) and *De donkere kamer van Damocles* (1958),\(^{39}\) reflect his breadth and talent. Recall that the data chosen for tests 1-3 are from a large 38-million-word corpus, containing many heterogeneous texts. Another reason for integrating literature into this dissertation pertains more to the eventual, marriage between traditional literature and linguistics studies. These two have often been at arms with each other, yet with the ever-increasing spread of the digital humanities, it has become apparent that both sides need to find a way to collaborate. Perhaps literary studies and linguistics can benefit from more cooperative objective tests. In linguistics, the use of literature has a long-standing tradition, but the opposite has not necessarily been the case. Finally, the choice to adapt literature was one of pure curiosity. If a linguist posits a reason for variation in grammar/lexicon and then fails to show it occurring in larger context, then the suggestion lacks deeper understanding of phenomenon in its entirety. Ideally, the final test would be spread out over the oeuvre of Hermans and many other authors, but the sheer amount of time involved in such an endeavor is staggering. Therefore, test four will observe split pronominals only in two longer works by Hermans.

\(^{39}\) The versions used in this study come from the 2010 anthology released by the Stichting Willem Frederik Hermans Instituut, see Hermans (2010).
8.1 Methodology

To begin, both novels were only available in print form. It was therefore necessary to scan both novels and render them machine readable. This was completed by scanning and then through making them machine readable via Adobe Acrobat. A major issue with creating machine readable texts lies within the process known as optical character recognition (hereafter OCR). OCR is currently not at 100% accuracy and requires a great deal of editing on the part of the researcher. As an example, the scans of both books were readable to the human eye with little or no difficulty, yet upon completing the OCR via Acrobat, numerous errors appeared in the texts. Some lines were scanned properly and still were illegible to the computer, making searching for the pronominal adverbs somewhat difficult.

Upon completion of the text clean up, the next step was to determine which type of pronominal adverb would be the target of this study. Tests one through three focus only on daarmee, and this was an acceptable limiter. With 38 million words at our disposal, the results were both copious and varying. However, when using only two novels, it became clear that the amount of daarmee extracts available was somewhat limited – seven instances in the split form and five in the unsplit in one novel, and one split and four unsplit in the other. The choice was then made to include all demonstrative pronominal adverbs that use daar as their first element. This choice was made over the other pronominal adverbs, er (personal), waar (interrogative and relative), and hier (the other option for demonstrative) as daar was the easiest to find and generally occurred more frequently in both forms. With er, nearly all of the extracts were split in the two novels. Hier appears too often in the text to distinguish easily from its separate locative adverbial form. All prepositions in conjunction with the demonstrative pronominal adverb daar were counted, not just met.
As with the previous tests, each unsplit form needed to first be checked to ensure its ability to split. If the pronominal adverb could not be split, then it was discarded. Factors determining whether it could be split or not were the same as above.

There is enough space to split the pronominal adverb:


The pronominal adverb was not assuming another function, i.e. daarvoor, daarbij, or daardoor, all having different functions unrelated to the pronominals under study:


(33) Daarbij fantaseer ik: - Volgens professor Nummedal bestond er geen twijfel dat de luchtfoto's hier waren (Hermans, 2010, p. 458).

(34) Daardoor is hij juist heterosexueel (Hermans, 2010, p. 702).

Once these examples were removed from the data, the next step was check the split forms. This step provided significantly fewer problems. Once all instances of daar were found, I simply read on to see if the example included a preposition or not.

The next level of processing includes marking for use spoken or narration context. Since this option was not available with the corpus data, it was utilized here. Each extract below is tagged for split or unsplit use in either narration or spoken context. For De donkere kamer van Damocles, I was able to find two narrative passages, one split and one unsplit, that facilitate understanding the hypotheses suggested. Since Damocles relies very heavily on spoken dialogue, two unsplit and two split forms appear here. Nooit meer slapen provided more narrative passages,
and as such I was able to obtain two unsplit and one split form. For spoken dialogue, there are one unsplit and two split forms. These numbers are not necessarily indicative of how often one form is used than the other, and that was not the goal of this aspect of test four, but by looking at use one notices that for *Nooit meer slapen*, the split form appears in narration 38% of the time, the unsplit 79%. For spoken passages, the split form occurs 59%, unsplit 21%. *De donkere kamer van Damocles* shows a more even spread. The split form is equal in spoken and narrative passages. The unsplit form appears 43% in narrative passages and 58% in spoken dialogue. We would like to reiterate that the purpose of marking for spoken or narrative passages was not to explicity test whether one form would appear more in type of passage or the other. Our goal here was to attempt to provide a balanced data set, with spoken and narrative passages receiving equal attention in both forms. Nevertheless, it is interesting that the two novels show different tendencies for the split or unsplit forms to co-occur within narrative or spoken passages.

For each extract, the exact verb was found. This includes looking for any idiomatic expressions or phrasal verbs, rather than just the infinitive itself. In doing this, I was able to determine if the use in either form can be attested to an idiomatic expression or to a somewhat novel construction with the verb in its "full", non-idiomatic form. This is important as it underlines the frequency tests performed in test two. Should the verb appear as an idiomatic expression, these require much less semantic parsing and are cognitively salient, therefore they will, by virtue of their frequency, require less attention. Should the verb be a novel construction, i.e. a construction that uses the verb in its "full" sense, not necessitating any further context to explicate its meaning, then this would require more cognitive energy on the part of the reader. Thus, test four also continues to use the frequency hypothesis to support the overarching

---

40 An English comparison would be similar to "The machine shreds paper" – full form, versus the idiomatic form "Kenny G shreds on the soprano sax". One easily imagines a machine physically shredding paper, but Kenny G is not physically shredding his soprano saxophone, but rather idiomatically shredding, i.e. playing with great intensity.
meaning difference. Namely, the unsplit form, collocates often with less-frequent verbs, thereby signaling a possible shift in discourse, requiring a more attentive reader/listener.

In order to implement the hypotheses suggested in tests two and three, this final test takes the data, after cleaning and marking, and extracts from both books pronominal adverbs that appear in both the split and unsplit form. These extracts are marked for narration or dialogue, whether the verb in the form is used idiomatically or in its "full" form, and if there is a shift in discourse being signaled. We determine a shift in discourse if the sentences following actively foreground or provide newer, more relevant information about the previous utterance or move the discourse in a different direction. Should the information following the form continue in the same vein, there is continuation of discourse, hence no signal for change. We also determine the semantic categories for each of the verbs in the 24 extracts. This was done via WordNet, as in test three. This reflects the findings that verbs of change should appear more with the unsplit form than the split, especially if a shift in discourse happens.

8.2 Implementation of Test Three on the Novels

Semantic category testing follows the same procedure outlined above. The results from the implementation of this test on the data lack the dynamism seen in the test on multiple verbs. As there are only 12 extractions used, one cannot expect a great variety or even to have the data here reflect the previous test. Nevertheless, Table 17 and Table 18 display how many of each semantic category appear and whether the extract was narration or dialogue.
What is apparent from the data is that there are far more stative verbs in dialogue than narration, and the values are the same for both forms. This occurs here because the verb *zijn* is also counted, whereas in the data in tests two and three, *zijn* was omitted entirely from the count. Ideally, the same would be possible here, yet on account of the limited data, all verbs needed to be included.
No values represented the change category in either form, yet because communication was the second highest ranking value for the unsplit, this is also reflected in the data. As mentioned above, the stative forms included *zijn*, so the data reflect a different trend in that regard.\textsuperscript{41} In the following section, I discuss the various passages from the book and show how each one either continues the line of discourse (split) or shows a possible shift (unsplit).

### 8.3 Semantic Integrality

The one underlying theoretical approach utilized in the final test is Yishai Tobin’s concept Semantic Integrality. Tobin purports that Semantic Integrality is "the perception of a concrete or an abstract entity within a continuous spatio-temporal-existential cline" (1996, p. 131). This approach is similar to Honselaar’s (1994) synthetic vs. analytic construction reading of the two forms, yet differs in that it can be extended to nearly any cognitive set in language. For example, Honselaar abstracts the split construction as (X-Y), forming one unified synthetic block of information; the unsplit is the separated, but contextually connected, analytic structure of ((X)-(Y)). In Semantic Integrality presents a similar situation. There are two abstracted constructions, namely construction 1: \((a + b = a + b)\) and construction 2 \((a + b = [ab])\). With the first construction, \((a + b = a + b)\), the concept of Semantic Integrality states that this is the natural reading, it is unmarked for semantic integrality. The two entities are discrete, perceived in a discontinuous spatio-temporal-existential cline. This abstraction is unmarked for Semantic Integrality. Construction 2 reflects different entities that are perceived as uniform and therefore marked for Semantic Integrality (Tobin, 1996).

As an example, Tobin analyzes the suggested synonymy between English *too* and *also* in terms of Semantic Integrality. Using this concept Tobin suggests that *too* is marked for Semantic

\textsuperscript{41} Four of the 12 extracts had *zijn*. Of these four, three of them appear in the unsplit form. In the previous tests, all instances of *zijn* were purposefully removed as it was difficult to separate full verbs from their auxiliary forms. Since this data set is much smaller, this distinction was much easier.
Integrality, being perceived as part of a single unit in the discourse. Also, on the other hand, is not marked for Semantic Integrality and therefore is not necessarily imagined as occupying the same spatio-temporal-existential cline; it must not be seen as a single unit. (Tobin, 1990). As is also the case with Tobin's analysis of Hebrew's negative existential particle *ein* – 'there is/there are no, not; does not exist; has/have not; not, no, else', the form unmarked for Semantic Integrality is also marked for focus (1996). In the case of the unsplit form, I also envision a marker of focus, as it is phonetically more dominant than the split form and also provides more information up front.\(^\text{42}\) Therefore, the unsplit form will be considered unmarked for Semantic Integrality.

The following examples of how Tobin uses Semantic Integrality come from his analysis of *Alice in Wonderland*, and are a source of Tobin's explanation. Examples that use *too*, Tobin states, emphasize "integrality of an addition" (1990). In example (35), Alice defends herself when the Dormouse accuses her of growing too rapidly (Tobin, 1990, pp. 160-161):

(35) "You've no right to grow here," said the Dormouse.

"Don't talk nonsense," said Alice more boldly: "you know you're growing too."

"Yes, but I grow at a more reasonable pace," said the Dormouse: "not in a ridiculous fashion."

Tobin uses this extract to show how Alice sees her growth as connected with that of the Dormouse. She implies unity between the growth that she and the Dormouse are undergoing. Nevertheless, the Dormouse retorts and does not accept her statement of unity. He does not see that they are growing similarly, but rather in different manners and thereby destroying Alice's implication of unity. To show how *also* is unmarked for Semantic Integrality, Tobin chooses the

\(^{42}\text{See reference to Bolinger in Figure 12.}\)
only two instances where *also* appears in the text. Example (36), originates from the story of Alice finding a baby, believing it to be a pig. In this example, Alice contemplates the possibility of the baby actually being a pig, and she describes the living object she holds.

(36) *The baby grunted again,* and Alice looked very anxiously into its face to see what was the matter with it. There could be no doubt that it had a very turn-up nose, *much more like a snout than a real nose; also its eyes were getting extremely small for a baby:* altogether Alice did not like the look of the thing at all. "But perhaps it was only sobbing," she thought and *looked into its eyes again, to see if there were any tears.*

*No, there were no tears.*

Here, Alice contemplates the true nature of the creature. She is unsure and describes the being, giving an inventory of its physical traits. Tobin believes that *also* used here can be replaced by *and,* further displaying how the description of the creature's turn-up nose and eyes do not particularly occupy the same internal continuous space. Were one to replace *also* with *too,* the clause initial position would not allow for the use, and the semi-colon creates a symbolic break between the two entities of its nose and eyes.

For this final test, Tobin's concept of Semantic Integrality provides the theoretical foundation. In the test three, I discuss how a semantic difference between the two forms, split and unsplit *daarmee,* is seen as a signal to the reader/listener of continuous discourse (split) or discontinuous shift in discourse (unsplit). Via Semantic Integrality, I suggest the unsplit form is unmarked for Semantic Integrality. The unsplit form does not necessarily occupy the same spatio-temporal-existential cline as the other elements. This was also exhibited in test three, where more verbs signaling change appeared. The split form is marked for Semantic Integrality.
Items in clauses with the split construction tend to show a continuous path of discourse and the reader should imagine the enclosed information as tying into the overarching discourse. We see this in the preponderance of static verbs, as evinced from test three. The following section will analyze individual extracts from the two Hermans novels listed above, and I aim to show how each split/unsplit construction fits into the concept of Semantic Integrality.

8.4 Extracts

With literature, it is much easier to catalogue change in discourse, as one has a continuous storyline upon which to base any analysis. In these two works, the situation is no different. Both novels provide strong storylines and develop as is expected in literature. In other words, these are not experimental works where one would be hard pressed to find a sense of continuity. Each extract below includes an analysis of what is happening and an explanation as to how the form, split or unsplit, can be understood as either maintaining or shifting the discourse. We use a total of 12 extracts. The extractions will proceed as follows: 6 passages from De donkere kamer van Damocles, followed by 6 from Nooit meer slapen. In the 6 passages, the sequence alternates from split to unsplit. We did this to provide a balance and to show that one must look beyond spoken or written forms as the reason for splitting. Discussions of each extract includes an explanation on how Semantic Integrality aids in understanding the continuous/discontinuous hypothesis put forth in test three. Background information is also provided to frame the story. Translations follow each extract.44

---

43 As an example, any test for Semantic Integrality in Joyce's Finnegans Wake would be neigh on impossible.
44 The English versions are taken from the following: Roy Edwards' 1958 translation of De donkere kamer van Damocles and Ina Rilke's 2006 translation of Nooit meer slapen.
8.5 De donkere kamer van Damocles

Summary

The first novel used for this test is Willem Friedrik Herman's 1958 novel De donkere kamer van Damokles. Because so much controversy surrounds this novel and the plethora of theories regarding its characters, in particular the protagonist, this summary will focus solely on the actions without further interpretation. Any interpretation or further information about the characters and actions is provided in the discussion of the subsequent extracts. For a compact, yet poignant, interpretation of the novel, I direct the reader to Janseen (1976).

The story follows Henri Osewoudt, a 20-something employee in his uncle's cigar shop. When Osewoudt was 12, his mother murdered his father in a fit of rage, and Osewoudt was thereupon sent to live with his uncle, Oom Bart. His mother was sent to a psychiatric clinic. Growing up, Osewoudt had few friends, was not particularly attractive, had thin hair, an effeminate voice, and shared a bed with his cousin, Ria, nine years his senior. When his mother was released from the psychiatric clinic, Osewoudt decided to marry his cousin, Ria, and take care of his mother. Osewoudt's life appears mundane, as he "had het gevoel of alles wat her had moeten worden gedaan, al gedaan was" (Hermans, 2010, p. 26).

When the Germans invade the Netherlands on May 10, 1940, Osewoudt's life changes. A Dutch lieutenant named Dorbeck appears and requests that Osewoudt develop photos for him. Dorbeck is sought for murdering Germans officers in Haarlem. Dorbeck has an uncanny physical resemblance to Osewoudt, yet is otherwise his polar opposite: strong, powerful voice, and dynamic. Days pass and Osewoudt hears nothing from Dorbeck, yet one day Osewoudt's doppelgänger reappears, hides his uniform, and borrows a jacket from Osewoudt only to disappear again. This sporadic interaction continues a while until Dorbeck finally disappears for

45 "[...] and he felt as if everything that had had to be done had already been done" (Hermans, 1958, p.18),
four years. Upon Dorbeck's return, Osewoudt is asked again to develop more photographs, yet Osewoudt does not know how to develop photographs nor what he is to do with them. These photographs later become the only clue he has to Dorbeck's existence. Unfortunately, the only photograph he develops that shows Dorbeck is ruined during development. His mother opens the door to the darkroom.

As time passes, Osewoudt finds himself more embroiled in the Dutch Resistance movement. He meets various members of a resistance society, is sent on missions to obtain information about German troop movement, as well as murder others. Osewoudt is eventually caught by the Germans and brought in for questioning. A prisoner, Roorda, mistakes Osewoudt for Dorbeck, he who killed the German officers. Osewoudt maintains his silence and is severely beaten. It is here that he meets Hauptsturmführer Ebernuss, the officer in charge of Osewoudt's interrogation. The two develop a somewhat amicable relationship and Ebernuss has Osewoudt released from the station and brought to a hospital. Later, Osewoudt is liberated from the hospital by Germans. He is informed that he is being freed from the hospital because they were, in fact, looking for his doppelgänger- Dorbeck. Osewoudt appears to be released from guilt.

Once again, Osewoudt meets with Ebernuss and they decide to flee together; Ebernuss wants to go AWOL. At a meeting of the resistance society, Dorbeck instructs Osewoudt to poison Ebernuss. Beforehand, however, Osewoudt requests a photo with Dorbeck. The photo is shot and shortly thereafter, Ebernuss is dead, Osewoudt takes off for Voorschoten (his hometown) and murders his wife Ria, as Dorbeck had informed Osewoudt that she turned him over to the Germans.

In 1945, shortly before the end of the war, Osewoudt is caught by the English, brought to Manchester, questioned and then released to the Netherlands to be imprisoned among Dutch
traitors. While in the prison, Osewoudt discovers that he may have been responsible for the deaths of many resistance fighters. His friendship with Ebernuß, the German interrogation officer, endangered the resistance, as he was bringing the enemy close to them. The Dutch do not give credence to Osewoudt's story about Dorbeck and believe him to be responsible for the deaths of not only the German officers, but the resistance fighters. Yet Osewoudt hopes for proof of Dorbeck's existence. The photo Osewoudt took of him and Dorbeck before killing Ebernuß must be developed. Upon developing the film with this photo, he notices that the image is gone. The only photo remaining is that of him with the German officer Ebernuß. Desperate, Osewoudt runs out of the building, trying to escape his distorted reality and is shot dead.

For each extract in this novel, I also use Janssen's summary of the work found in the Synthese series. This guide helps demarcate episodic boundaries. Since de donkere kamer van Damocles does not have chapter titles, one is left to his own devices to draw episodic boundaries. At times, these episodes are located on the next page, or within the page itself. Nevertheless, it is not entirely clear where each episode begins and ends. Janssen makes note of this in his summary, "They [the episodes] all have varying lengths. Often segments are introduced inside of the episodes via extra space between the lines; these segments are neglected in my summary. The episodic numeration is my choice" (Janssen, 1976, p. 16).46 We notice that the elements discussed in the unsplit sections often correlate to major episodic boundaries that are reflected in Janssen's summary. Split forms, by contrast, do not figure into the episodic boundaries suggested by Janssen. Events in the split form passages may be mentioned in the summary, but they are not the focus of the episode.

46 "Ze zijn onderling zeer ongelijk van lengte. Vaak zijn binnen de episoden weer segmenten aangebracht die door een witregel aangegeven zijn; deze segmenten worden in mijn overzicht verwaarloosd. De nummering van de episoden is van mijn hand".

89
Extract 1
De donkere kamer van Damocles - p. 187
Form: split daarna
Narration

Niemand heeft mij gevraagd hoe ik aan die papieren gekomen ben, dacht Osewoudt terwijl hij op de brancard door de gang zweefde. Waarom heeft Ebernuss daar niet naar gevraagd? Waarom hebben zij mij kapot geslagen om te weten wat ik met Elkan heb uit te staan, die ik nooit heb gezien? Waarom laten ze mij zogenaamd herkennen door die zogenaamde Roorda? Wat willen ze daarmee bereiken? En geen woord over mijn valse papieren!

"No one has asked me how I came by those papers," thought Osewoudt, as he floated down the corridor on the stretcher. Why didn't Ebernuss ask about them? Why have they beaten me to pulp in order to find out what my connection is with Elkan, whom I've never seen in my life? Why do they have me 'recognized' by that 'Roorda'? What do they think to achieve by that? And not a word about my forged papers! (Hermans, 1958, p. 175).

In this first extract, the concept of the documents appears both before and after the split pronominal adverb, signaling a continuous idea through the storyline. Also, the split pronominal adverb begins a series of questions, highlighting the narrator's confusion while being transported on a stretcher. With regards to the story, Osewoudt, the protagonist, is being removed from his

cell. Ebernuss, the SS officer in charge of interrogating Osewoudt, has found falsified documents with the name "Filip van Druten", one of many aliases used by Osewoudt to hide his identity from the SS. Since, however, Ebernuss knows who he is, Osewoudt becomes confused and concerned as to why they were not asking about his falsified documents, but rather about people he has never met. Elkan is another character who supposedly participated in the assassination of an SS officer in the Kleine Houtstraat and looks very similar to Osewoudt. It never becomes clear who Elkan really is, and it is possible that Elkan is indeed another person, or one of the many instances of doppelgängers throughout the novel. For this extract, it is apparent that the discourse is continuing through and the split form signals no change.

With regards to Semantic Integrality, the split form here highlights a desire to signal a continuous discussion about the documents. The focus is not on what Ebernuss asked, but that he asked. The documents, and the repercussions of having them, continue to affect Osewoudt and his situation. In this extract, Hermans may not want to lead the reader to think that inquiring about the documents necessitates focus, but instead Osewoudt's constant asking "waarom"- why - appears to be the focus here.

Extract 2

De donkere kamer van Damocles - p. 332

Form: unsplit daaruit

Spoken

Ze hebben Osewoudt dus naar het ziekenhuis Zuidwal laten brengen. Daaruit is hij dezelfde dag nog door ons zogenaamd bevrijd. War eine tolle Geschichte!
That had Osewoudt taken to Zuidwal Hospital. From there, he was "rescued" by us, on
the same day. War eine tolle Geschichte. [It was a lot of fun! – my translation]

(Hermans, 1958, p. 312)

In this passage, the unsplit pronominal adverb functions as a way of showing a major
shift in the story. Osewoudt had been previously under arrest and was being interrogated by the
Germans. We see this in the previous excerpt. The speaker is Malknecht, a German secretary
who had been taking notes during the interrogation. Ebernuss had arranged that Osewoudt be
"liberated" from his cell by two Dutch agents working for the Nazis who were posing as
members of the Dutch resistance. Up until this point, Osewoudt thought that he had been
liberated by Dorbeck. Upon revealing that Osewoudt had, in fact, been liberated by Nazi
sympathizers, a major shift in the story becomes apparent. His liberation took place in order to
have Osewoudt lead the Nazis to other members of the resistance. Osewoudt does not want to
believe the reality of his supposed liberation. In his mind, Dorbeck organized the liberation.
Therefore, a clear shift in discourse occurs in this extract. What was once considered truth
changes into lies. A major change takes place in the storyline and the protagonist Osewoudt is
forced to confront different realities.

Semantic integrality does not come into play here. Since there is an immense break in
assumed realities, suggesting, via a split form, that the two ideas should be seen as unified
somehow (Osewoudt's assumption of a Dorbeck-orchestrated liberation vs. reality), is misleading.
The unsplit form here signals to the reader that what occurs within the unsplit construction stands
in stark contrast to the previous understanding of reality. It is unmarked for Semantic Integrality.
Extract 3

De donkere kamer van Damocles - p. 243

Form: split daarmee

Spoken

- Tussen haakjes, zei Ebernuss, je moet mij ook maar Waldemar noemen. Wij kennen elkaar al zo lang, maanden en maanden ... Maar dat is de enige reden niet. Je hebt niemand meer in de wereld, denk eraan. Je moeder is dood, je oom is dood, je vriendin Elly Sprenkelbach Meijer is dood, Labare is dood, je vriendje Robbie is dood of in een concentratiekamp waar hij in geen geval levend uitkomt. Maar begrijp mij goed: ook ik sta alleen. Al mijn vrienden zijn gevallen aan het front. Mijn moeder ligt onder het puin van haar huis in Frankfort. Wie zal mij helpen als de Amerikanen komen?

Osewoudt ging rechtopzitten, keek Ebernuss aan met een vuile grijns en zei: - Niemand denk ik. Dan zijn de rollen omgekeerd!

- En wat dan nog als de rollen omgekeerd worden? Krijg jij daar je moeder en je vrienden mee terug? Nou? Wat zeg je?

Osewoudt vouwde de wijsvinger van zijn rechterhand en beet op de knokkel.

Ebernuss legde zijn hand op zijn schouder.

'By the by,' said Ebernuss, 'you ought to call me Waldemar, too. We've known each other for so long now, months and months... but that's not the only reason. You've no one left in the world, mind. Your mother's dead, your uncle's dead, your other girl friend Ellie Sprenkelbach Meyer is dead, Labare is dead, your little pal Robbie is dead, or at
any rate he's in a concentration camp which he'll never leave alive. But – get this straight: I stand alone, too. All my friends have fallen at the front. My mother lies under the rubble of our house in Frankfurt. Who's going to help me when the Americans come?'

Osewoudt sat up, looked at Ebernuss with a dirty grin, and said: 'Nobody, I think. Then the boot's on the other foot!'

'And what if the boot is on the other foot? Will that give your mother and your friends back to you? Eh? What do you say?'

Osewoudt crooked the index finger of his right hand, and bit the knuckle. (Hermans, 1958, p. 228).

This excerpt finds Osewoudt back in his cell, being interrogated by Ebernuss. The sentence with the split pronominal adverb appears much later in this except, yet it is important to observe the previous dialogue. Ebernuss is trying to make a point that both he and Osewoudt has no one left. Ebernuss lists the names of those close to Osewoudt who are dead or are currently in captivity. Then Osewoudt switches the topic a bit, focusing on the possible liberation via the Americans, but this idea is promptly shot down. Ebernuss reminds Osewoudt that even if the Americans came, and the tables were turned, nothing could be done to bring back those he has lost. Here the split form continues with the discourse mentioned above. Ebernuss wants to remind Osewoudt of the reality, he wants to pull him back into the topic of being alone. For this reason, a split pronominal adverb is used, not signaling any shift from the original topic.

Ebernuss, by using the split form, signals that the ideas exist in a continuous space. They are all related in the sense that the deaths of those Osewoudt has loved and a shift in fate for the
Germans vs. the Americans are interconnected. Even if the Germans fail at their objective in the war, this will not bring back the dead. Hence, an instance of the split form appears marked for Semantic Integrality.

As the discussion continues, Ebernuss attempts to console Osewoudt, informing him that a Jewish girl Osewoudt knows could be freed, but simultaneously he also wants Osewoudt to admit to having met with Roorda (one of the men responsible in the killing of an SS officer). Ebernuss's goal here is to play with Osewoudt's emotions and thereby obtain a confession. All of what Ebernuss says attempts to unify the reality he wants to create for Osewoudt.

Extract 4

De donkere kamer van Damocles - p. 361

Form: unsplit daarop

Narration

The dentist laid his case open on the grass. From it he took a big chart of yellow cardboard. On that, the teeth of the lower and upper jaws of the human mouth were shown diagrammatically, in two horseshoe-shaped semicircles opposite each other. Here and there, various marks and signs had been made on the diagrams. On the left, all kinds of details had been recorded about the patient concerned: surname, Christian name, date of birth, the dates on which he had been treated. The dentist happened to cover these particulars involuntarily with his left hand, but nevertheless Osewoudt could read: Jagtman, Egbert, 3 December 1916. No less than three addresses, two of which had been crossed out (Hermans, 1958, pp. 338-339).

Upon developing photographs given to him by Dorbeck, Dorbeck instructs Osewoudt to send these photos to a particular Jagtman. The photos he was to send to Jagtman were useless, as Osewoudt's mother enters the darkroom, thereby destroying all photos. In one of these photos Dorbeck appears in front of a house that later turns out to be related to Jagtman. Jagtman is also one of the only real personal connections Osewoudt has to the illusive Dorbeck. Much later in the story, Osewoudt discovers that Jagtman's family dies when a plane crashes into their third-floor apartment. In addition, Jagtman has turned himself into the German police and was brought to a concentration camp in Oldenburg. Many believe that Jagtman and Dorbeck are the same person, nevertheless Osewoudt insists that they are two separate people. Jagtman is Osewoudt's key to discovering the true identity of Dorbeck. The situation then takes a sharp turn when Osewoudt discovers that the dentist, in the passage above, was Jagtman's dentist. The dentist states that Osewoudt seems to resemble Jagtman somewhat and has information that could identify the body of Osewoudt, a body that Osewoudt thinks could be Dorbeck's.
We see how the narration utilizes the unsplit form, signaling a major development in the story. The clause could have been written split as such:

(37) Kiezen en tanden van onder- en bovenkaak waren daar schematisch op aangegeven [...] 

Hermans opts, however, for the unsplit form. This major shift in development also fits into the hypothesis provided in test three. It is now possible to identify the corpse in the Oldenburg concentration camp. It could be Jagtman, who may possibly turn out to be Dorbeck. Were the clause to have been written as split, the reader may have taken less notice of the significance of the event. Also, if it were split, Tobin's Semantic Integrality would signal that the action of opening the suitcase, laying it in the grass, pulling out the card and all such further actions occupy a continuity of time. However, time seems to shift abruptly when Osewoudt recognizes that the card has vital information that could lead to finding Dorbeck. This instance of the unsplit exhibits just how powerful it can be. Finally, there is a connection, and it may be possible to identify the one person, outside of Osewoudt, who had any connection to Dorbeck.

Extract 5
De donkere kamer van Damocles - p. 246
Form: split daarop
Spoken

al een kwartier eerder was gekomen. - Anders heb je niets te doen. Het is heel
eenvoudig. Ik ga niet met je mee, niemand van ons gaat mee, **daar kun je op rekenen.**
Je hoeft niets anders te doen dan een borrel te drinken op die zolder.

'Well then... now you must think of Mariane again. I'll make you the following
proposition. Within the next few days, we'll take you to that dive. You'll go in, you'll
say a few words to Moorlag. You must take good note of how you're received. To give
an example—they may address you as Dorbeck. Or they may say: "Hey, what's this?
We thought you arrived a quarter of an hour ago." You don't have to do anything else.
It's very simple. I shan't go with you, not one of us will go with you, you can reckon on
that. You don't need to do anything but have a drink in that attic" (Hermans, 1958, p.
231).

Ebernuss is planning to sabotage a meeting of clandestine underground heroes of the
Dutch resistance movement. During this meeting Dorbeck is to appear. For this purpose,
Ebernuss is sending Osewoudt into an attic where the group is meeting and he wants Osewoudt
to uncover Dorbeck. Osewoudt is to find the organizer of these meetings, Moorlag, who
supposedly had lived at his house for a while. Osewoudt then needs to have a short conversation
with him and leading to Moorlag's capture. Osewoudt will be on his own.

In this passage, the use of the split form appears with an idiomatic expression: **op iets
rekenen** – to count/rely on something. As stated in test two, idiomatic expressions often occur in
the split form, as they are easier to parse for the reader/listener. The use of the split here simply
signals to the listener that the fact whereupon he can rely is something already known to him.
And since he knows this fact already, it's occupying the same existential plan, hence a signal of Semantic Integrality. This supports the discourse from before and does not provide any new information. Osewoudt will not receive any backup; he is entirely on his own. Therefore, the split form continues with the previous discourse, not adding any new information.

Extract 6
De donkere kamer van Damocles – p. 119
Form: unsplit daarover
Spoken

'I don't know anything! You're the one who's telling me that the Germans send mentally deranged people and recidivists to the concentration camps without a trial! I've said nothing about that! All I said was that I don't understand why an old woman who did violence to her husband ten years ago, but who's been before the courts and was in an institution for a time, and never hurt a fly afterwards – I tell you I can't understand how
she could be arrested again, just like that! And what about Ria? What has Ria done?'

(Hermans, 1958, pp. 105-106).

In this extract, Uncle Bart and Osewoudt are discussing the current situation of the German occupation. Osewoudt has just explained to his uncle that the Germans are packing up criminals and the mentally ill and shipping them off to concentration camps. What Uncle Bart cannot understand is why the Germans simply arrested Osewoudt's mother (who killed her husband in a fit of rage ten years prior) and Osewoudt's wife, Ria. Osewoudt knows nothing.

Whereas it is true that Osewoudt's mother did kill her husband, she served her time. Ria, Osewoudt's wife, was apparently innocent, but Osewoudt does not seem the slightest bit concerned for her safety. Uncle Bart is shocked at Osewoudt's apparent lack of interest in his mother's and wife's well-being and is becoming angrier with Osewoudt. Still, Uncle Bart insists that he knows nothing about the arrests, for if he were to know that would suggest that he is listening to British radio- a crime punishable by 2 years in prison. Uncle Bart's denial of any knowledge regarding the arrests is underlined by the use of an unsplit pronominal adverb. He is stressing that he knows nothing ABOUT THAT. The unsplit pronominal adverb, in terms of Semantic Integrality, also shows a break in a break in an assumed continuous spatio-temporal-existential plane. All of what Uncle Bart states beforehand stresses that Osewoudt is to blame for what Bart knows. The unsplit shows the break in that continuum, now the reader moves to what Uncle Bart says himself, what he believes.

Since he is aware that a prison sentence under the Germans basically equates to a death sentence, he is very concerned about informing Osewoudt. The sentence following the unsplit construction shifts the focus of the discussion to what Uncle Bart did say: "Ik heb alleen gezegd
The reader maintains a similar trajectory, but is moved away from the initial path— that of Osewoudt's statement. Uncle Bart wants to maintain his distance and safety, and accomplishes this with the use of the unsplit pronominal adverb.

8.6 Nooit meer slapen

Summary

The following summary shall make the reader aware of the more poignant scenes in the novel and does not attempt to provide a detailed description. Character analysis, where appropriate, appears in the subsequent extracts. As with the summary above, this is to serve the reader with a brief synopsis of the story before reading the extracts and has benefited from the Synthese series, in particular den Boef (1984).

In contrast to De donkere kamer van Damocles, the second novel, Nooit meer slapen, does not move in and out of characters, nor does it contain much interaction with multiple characters. Hermans brings the reader into the mind of the protagonist, Alfred Issendorf, a Dutch student of geology. Alfred flies to Norway in order to conduct his dissertation research in the desolate northern area of Finnmarken. His goal is to find any traces of meteorites and thereby support his doctoral advisor's theory that these created the craters one finds in the region.

His story begins with a difficult meeting between him and Prof. Ørnulf Nummedal. Prof. Nummedal and Alfred have difficulties communicating as the professor prefers to speak in German, whereas Alfred would rather speak English. Nevertheless, the reader follows the two characters on a day trip through Oslo, during which Prof. Nummedal's apprehensions regarding Dutch geologists becomes apparent. From here, a sense of despair arises in Alfred and the reader senses his slow academic demise. In addition to both the communication difficulties and Prof. Nummedal's critical stance toward Dutch geologists, the reader also discovers that Alfred's
advisor, Prof. Sibbelee once confronted Prof. Nummedal during a conference many years ago, an act that was not soon forgotten.

Alfred flies to Oslo not to meet with Prof. Nummedal, but to procure aerial photographs of the region of Finnmark he will study. After spending a few days wandering Oslo with Nummedal and discussing geology, *gravlaks* (a Norwegian fish specialty), and generally wasting time, Alfred is disappointed to discover that the photographs he so desperately needs for his research have been sent to the north, to Trondheim. Upon flying north, Alfred tries to meet with director Hvalbiff, but is told that he is out of the office and that the aerial photographs are not available.

Without his much-needed aerial photographs, Alfred travels further north to Tromsø and subsequently to Alta, where he meets his friend, Arne Jordal. There, they continue to travel to Skoganvarre, where they encounter the third traveler, Qvigstad. It is here that Alfred begins to face the ordeals of life in the desolate wasteland of Finnmarken. As they continue, they meet up with a fourth traveler, Mikkelsen. Mikkelsen is a student of Nummedal and Alfred begins to sense that this new traveler may be out to sabotage his research. In a turn of events, Alfred discovers that Mikkelsen has the aerial photographs, and Qvigstad finds a way to allow Alfred to observe the photographs. Upon investigating the aerials, Alfred discovers that there is nothing to give him any inkling of hope that the theory has any ground, but he is determined to continue the expedition.

Qvigstad and Mikkelsen depart on their own, leaving Arne and Alfred. Alfred decides to climb Vuorje mountain to gain a better view of where Mikkelsen and Qvigstad may be heading. Upon scaling the mountain, his view is hindered by heavy cloud cover. Disappointed, Alfred

---

47 Alfred later learns that Hvalbiff means "whale meat", a possible cause for his difficulty in obtaining the aerial photographs.
returns, looking for Arne, yet finding only a tripod. Following the direction in which the tripod was aimed, Alfred discovers Arne, dead on the ground. This is the turning point in the novel, as Alfred decides to return and give up his expedition. While flying back to Holland Alfred reads that a loud explosion he heard while still in the north was, in fact, a meteorite. Alfred's mother and sister meet him at the airport. He is depressed, having returned without any proof of his hypothesis. It is here that his mother presents him with two cufflinks, inlaid with meteorite stones. Alfred returns home, not having fulfilled his destiny, in his eyes, and the reader is left to ponder his future.

All English passages for Nooit meer slapen come from Ina Rilke's 2006 translation.

Extract 7
Nooit meer slapen – p. 580
Form: split daarover
Spoken

Over de meeste scheppers wordt altijd weinig gepraat. Over stuntgeneraals, politici en andere bedriegers, daar wordt veel over gepraat, niet over scheppers.

‘Creative minds never get much of a mention. It's warlords, politicians and other confidence tricksters who get talked about all the time, not people with creative minds’ (Hermans, 2006, p. 175).

It's morning and the rain has begun to fall, and it has also started to seep through the tent. Alfred, the protagonist, and his travel companion, Arne, are inside. Unable to sleep, Alfred hears
the rain falling on the tent, contemplates flies healing his wound and then decides to wake up Arne. They start discussing the fate of intellectuals and creators. Lamenting the lack of praise showered on those who create, they discuss the dreadful reality of having to teach or work in a factory. How horrible would it be to have a talent that is of little use to anyone? To what purpose does the Greek teacher use his talents to create instead of just passing on a language that few will appreciate? How can the factory worker create when all he does is the same thing? This discussion continues for around a page.

Using the split form here allows the two of them to advance the discussion in the same path. The ensuing conversation is semantically integrated into the previous one. Tobin's Semantic Integrality, once again, shows how the split form presents a continuum. Alfred provides no signal to Arne that the topic is changing. By splitting the pronominal adverb, Alfred suggests that everyone else but the creators are glorified and Alfred expects Arne to provide more material for this conversation.

Extract 8

Nooit meer slapen – p. 414-15

Form: unsplit daarover

Spoken

- Ik ken professor Sibbelee al vele jaren. Wanneer zal ik hem het eerst ontmoet hebben?
Dat moet nog voor de oorlog geweest zijn, op het congres in Tokio. Ja. Het jaar waarin ik mijn, ik mag wel zeggen inmiddels klassiek geworden inzichten heb voorgedragen over de milonietzone in Värmland en de voortzetting daarvan in Noorwegen. Vielleicht kennen Sie die kleine Arbeit!
Hij zwijgt een tel, maar niet lang genoeg om me tot de bekentenis te dwingen dat ik genoemd werkstukje niet ken. Opgeruimd vervolgt hij zijn verhaal.

- **Sibbelee is toen met mij daarover in debat getreden.** Het is er warmpijes toegegaan.
In geen enkel opzicht kon hij zich bij mijn opvattingen aansluiten. Stelt u zich voor!
Wat een toestand! Sibbelee is dertig jaar jonger dan ik en hij was toen nog heel, heel erg jong. Geestdrift van de jeugd.

'I have known Professor Sibbelee for many years. Let me see, when did I first meet him?
It must have been before the war, at the conference in Tokyo. Yes. The year I presented my paper – which has become a classic, if I may say so – on the milonite zone in Värmland and its expansion into Norway. Vielleicht kennen Sie die kleine Arbet?'
He pauses for a moment, but not long enough to compel me to confess my ignorance of the said opus. Then, brightly, he continues.

'Sibbelee opened a debate about it at the time. Things got quite heated. He could not agree with a single argument I put forward. Can you imagine? Such a to-do! Sibbelee is thirty years younger than me and in those days he was very young indeed, very young.
The passion of youth! (Hermans, 2006, p. 4)

This passage is in the beginning of the story where Alfred meets Prof. Nummedal.

Alfred's dissertation advisor, Prof. Sibbelee, has suggested Alfred contact Prof. Nummedal in order to obtain much needed aerial photographs of the region he is about to research. Sadly, the 84-year-old Prof. Nummedal does not know of any communication with Prof. Sibbelee regarding the aerial photographs and then continues to recount a story of how, thirty years prior, Sibbelee
addressed him at a conference. The discussion starts harmlessly. Nummedal underlines his long-standing familiarity with Sibbelee. The reader has no idea about their relationship. Nummedal speaks of a conference, and he iterates, somewhat jokingly, "Vielleicht kennen Sie die kleine Arbeit" – 'Perhaps you know of the little piece of work' [Translation mine]. It is also important to note how he switches to German here, as this signals a turn for the worse, placing Alfred at a slight linguistic disadvantage.

This change is made clear by the use of the unsplit form. Once again, Hermans shifts the focus. Prof. Nummedal talks of his research, brags almost, and asks if Alfred knows of it. Without any hesitation Nummedal then shifts the focus from himself to the "warmmpjes" discussion he had with Sibbelee. Nummedal does not see the debate as "warmmpjes" in the slightest. As he continues to speak of the "geestdrift" of the youth, the reader notices just how embittered the 84-year-old Nummedal remains. Using the unsplit form here signals a disconnect with the previous statement, showing no semantic integrality. The reader quickly sees how dire the situation is.

This is significant as it sets the stage for Alfred's further collaboration with Nummedal. Later, Nummedal belittles the work of Alfred's advisor, thus making it even more so apparent that the aerial photographs Alfred so badly needs, will probably not be available to him. The unsplit form here signals a break from an initial, cordial greeting to one of distance and slight animosity. Alfred's situation only worsens from here out.

Extract 9

Nooit meer slapen – p. 596

Form: unsplit daarmee

Narration
Ik probeer het dons in mijn natte slaapzak door de bekleding heen zoveel mogelijk los te plukken. Misschien valt de schade nog mee.

**Terwijl ik daarmee bezig ben, valt mijn oog op Mikkelsen.** Hij ligt op zijn buik voor hun tent en kijkt... Hij kijkt door een stereoscoop. Hij heeft een stuk plastic op de grond gelegd en daarap liggen foto's die hij met twee tegelijk door zijn stereoscoop bestudeert. Wat zijn dat voor foto's? In de zwarte randen lijken witte punaises afgedrukt, maar dat zijn de meegefotografeerde klok en hoogtemeter. Luchtfoto's!

I take my sleeping bag and try to separate the lumps in the down through the fabric. Perhaps it won't be ruined after all.

While I'm busy doing this, my eye falls on Mikkelsen. He is lying stomach down in front of their tent...doing what? Peering through a stereoscope. Laid out on a sheet of plastic before him is a batch of photographs, which he studies two at a time through the instrument. What is in those photos? I can make out pale blotches in the black borders, could be the imprint of drawing pins. But the could also be part of the image – that is, the faces of clocks and altimeters. Aerial photographs! (Hermans, 2006, p. 191)

As stated earlier, Alfred is dire need of aerial photographs to better understand the area he is studying. Nummedal was incapable of providing them and something told Alfred that this inability may have had something to do with the disagreemnt his advisor had with Nummedal many years prior.

The scene begins relatively banal. The group is setting up camp, Mikkelsen is playing around with reindeer antlers he has found and all are trying to dry out their clothes. The reader is
focused in on Alfred, his thoughts about the pain from a previous accident and the astounding emptiness of the terrain. Initially, all seems to be void of emotion, even the swollen pain in Alfred's leg is no match for his lack of sleep. However, while Alfred is busy plucking feathers, something catches his eye. Mikkelsen has the aerial photographs he needs so badly. In this instant, a major shift occurs in the story. Alfred had been apprehensive of Mikkelsen before, finding him somewhat of a nuisance, but it was not until this turning point that Alfred's rage becomes apparent. Prior to this revelation, he states:

(38) Geen mens weet waartoe hij in staat is, voor hij alles geprobeerd heeft. (Hermans, 2010, p. 596)

There's no knowing how much you can take until you've tried everything (Hermans, 2006, p. 191).

The reader is pushed into Alfred's world of pain, sleeplessness, desperation. He sees the photographs and knows that something has gone awry.

The unsplit form here creates an orthographic boundary between the previous actions and the new revelation. In many instances in the data *met iets bezig zijn* regularly appeared in the split instead of unsplit form. This break brings the reader's attention to the imminent appearance of something significant. Indeed, it is significant. Alfred now knows that he had been tricked and Professor Nummedal was the one playing tricks on him. For this extract one clearly sees how the unsplit form is being used to signal a major shift in the discourse. In terms of Semantic Integrality, this unsplit clause is unmarked. The events prior to seeing the aerial maps may occupy a linear space with plucking down feathers from a sleeping bag, yet, the mere reality that these maps appear, breaks any continuous spatio-temporal relationship. Alfred shifts his attention,
his reality and pain to Mikkelsen. The reader's attention is warned with the unsplit form that something pertinent will occur, and indeed it does.

Extract 10

Nooit meer slapen – p. 598

Form: split daarvoor

Narration

Nu heeft Sibbelee iets nodig van Nummedal. Natuurlijk is het onder professoren niet gebruikelijk dergelijke verzoeken vlakaf te weigeren. **Bovendien is Nummedal daar veel te geraffineerd voor.** Hij pakt het anders aan. Hij ontbiedt zijn leerling Mikkelsen bij zich, suggereert dat Mikkelsen wel eens een interessant onderzoekje ter hand zou mogen nemen - *mijn onderzoek.*

Sibbelee needs a favour from Nummedal. It would obviously be bad form for a professor flatly to run down the request of a colleague, but Nummedal is too crafty for that anyway. More devious. He summons his pupil Mikkelsen and proposes an interesting little research project for him to undertake – *my research* (Hermans, 2006, p. 193).

From the previous extracts, the tension exists between Professors Nummedal and Sibbelee manifests itself. Also, an immediate problem appears between Alfred and Mikkelsen, Nummedal's pupil. Tensions and problems continue. Alfred is underway with Nummedal's pupil, Mikkelsen. Sibbelee has requested that Nummedal provide Alfred with aerial photographs, and
yet Nummedal does not provide them, as he gave them to his pupil, Mikkelsen. Knowing full well that Nummedal appears to be incapable of letting go of the disagreement with Sibbelee thirty years’ prior, Alfred begins to fear that he will become the target of Nummedal’s retribution. Instead of simply saying that he will refuse to help Alfred, Nummedal suggests Mikkelsen go along and thereby sabotage Alfred’s research. This appears to be the case. Mikkelsen has the aerial photographs Alfred so desperately needs.

The split construction appears mid-paragraph. Alfred contemplates how some people deal with petty arguments and then tries to reassure himself that Nummedal would not react thus. Nevertheless, he knows that some sort of retribution is imminent. The discourse continues along the path set out originally. In other words, it continues to occupy the same spatio-temporal-existential cline, being marked for Semantic Integrality. He continues to explain how Nummedal is too slick to do otherwise.

Alfred is aware that professors do not simply deny requests for help, but are rather craftier in their ways. More information related to the previous statements tags along, and it does not waver from the preset path. There are furthermore no other signals in the clause that would suggest any shift—such as a contrastive conjunction like maar (but). The passage might have been rendered:

(39) Maar daarvoor is Nummedal veel te geraffineerd.

But for is Nummedal much too crafty.

'But that's something for which Nummedal is much too crafty.'

Were this to be the case, one would understand that Alfred may have expected such a petty response, but is surprised at how Nummedal behaves. Still, Hermans opts for the split form and thereby shows that Alfred has not shifted in his thoughts. As the story continues, Alfred’s rage
toward Nummedal increases, referring to him internally as "beschimmelde vlees" (moldy flesh) and calls him a "stomme varken" (stupid hog). Hence, the split form exhibits how the reader is held to the discourse in the passage and nothing entirely new is presented.

Extract 11
Nooit meer slapen – p. 507

Form: split daaraan

Spoken

- **Ik ben ambieus, daar kan ik niets aan doen**, zelfs al weet ik waar het van komt.

Mijn vader was een veelbelovend botanicus, maar hij is verongelukt, net toen ik zeven jaar werd. Hij is in een spleet gevallen, in Zwitserland. Een paar dagen nadat wij het overlijdensbericht gekregen hadden, kwam er een brief dat mijn vader tot professor was benoemd. De sprekers aan zijn graf wisten niet precies of ze het over professor Issendorf moesten hebben, of over meneer Issendorf. Mijn moeder heeft mij opgevoed in het denkbeeld dat ik de carrière die hij niet heeft kunnen afmaken, moet voltooien.

'I'm ambitious. I can't help it, even though I know where I got it from. My father was a promising botanist when he was killed in an accident, just before my seventh birthday. He fell into a crevasse, in Switzerland. A few days after we heard of his death, a letter arrived saying that my father had been given a professorship. The speakers at his funeral weren't sure whether to refer to him as Professor Issendorf or as Mr Issendorf. My mother brought me up to believe I was destined to make up for his broken career by being successful in my own (Hermans, 2006, p. 99).
Alfred and his guide, Arne, are talking about Alfred’s theory regarding meteorite craters in the landscape. Arne finds the hypothesis somewhat shocking, but listens. When Alfred tells him how he gets heart palpitations thinking about the possibility of finding meteorites, Arne somewhat jokingly states that he will do his best not to think about it. Up until the point where this extract begins the reader knows that Alfred’s father is dead, but how he dies is not known. The reader is also clued into expectations placed on Alfred during childhood. His undying passion for his research stems from both an internal drive to excel and also his mother’s desire for him to follow in his father’s footsteps (a desire that later proves nearly fatal).

Appearing near the beginning of this particular element of discourse, the split pronominal adverb allows the reader/Arne to be clued into how Alfred plans to continue the discussion. Alfred is ambitious, and his ambition cannot be changed for particular reasons. Alfred wants to signal to Arne that he is incapable of altering his personality in this regard. He then continues to explain to him why that is. In addition to wanting to signal this continuation, the expression *niets aan iets kunnen doen* also tends to appear more often with the split than the unsplit form. 48 This falls in line with the hypothesis put forth in test three, whereby idiomatic expressions or fossilized constructions show a tendency to appear more with the split form than the unsplit. In fact, the expression with the split pronominal adverb may as well be omitted from the conversation as his inability to change this ambition is later explained.

---

48 A Google search for "daar kan ik niets aan doen" provided over 20,000 results, whereas its unsplit counterpart "daaraan kan ik niets doen" shows only 20.
Wanhopig probeer ik boven op de stenen te klimmen - mijn knie! mijn knie! Nergens vind ik houvast. Een andere, kleinere steen aanslepen om als trapje te gebruiken? Maar iedere steen die daarvoor groot genoeg is, is te groot dan dat een mens hem kan optillen. Ik neem een aanloop, werp mij tegen de laagste van de drie stenen aan. Mijn armen e roverheen. Kan met mijn vingers de achterkant bereiken, trek. Voor mijn part rolt de steen om als ik er zo aan hang. O! Ik word zo lief verpletterd! Belachelijke hoop natuurlijk, hij weegt minstens drie ton.

In desperation, I try climbing on top of the rocks - my knee! my knee! The sides are too straight. Should I try putting a smaller stone against the base for me to get up on? But for a stone to serve as a step it has to be quite big - too big for me to lift. I take a run-up and hurl myself at the lowest of the three blocks, slap my arms over the top, hook my fingers round the far edge and pull. Hanging on like this could make the whole thing topple backwards. Oh, to be crushed to death and be done with it all! No such luck—it must weigh three tons (Hermans, 2006, p. 231).

Alfred is now on his own and trying to find his way through the vast emptiness of Northern Scandinavia. Arne is lost and Alfred needs to find his compass that just fell between
the cracks of three enormous rocks. The compass was a gift from his younger sister, who was afraid that he would become lost. His sister's fear has come true. The compass has fallen and it is unclear whether he will be able to find it in the darkness between the rocks. He is not able to light a match to see if it is anywhere nearby, as all of his matches are completely wet. Nowhere can he find a stick long enough to plunge into the rocks. It appears that the compass will be lost forever.

Every stone is too big; Alfred has arrived at a point of no return. How is he to find his way through this vast emptiness without a compass? The unsplit form here signals yet another major shift in the narrative where the reader realizes that Alfred feels completely hopeless, "O! Ik word net so lief verplettered!" He'd rather be crushed to death by the stones than to have to continue on in this lost landscape. The size of the rocks is not clear until after the unsplit form. The stones are "te groot dan dat een mens hem kan optillen" – too large that a human could lift it. In this final example, the unsplit form exhibits another shift in the discourse. The situation goes from bad to worse, and Alfred is in dire need of direction, both via his compass and mentally. His travel companion, Arne, is also lost (and as he later discovers, dead). So, without a compass and without a fellow traveler, Alfred continues alone and desperate.

8.7 Review

The 12 extracts above, from both De donkere kamer van Damokles and Nooit meer slapen, have been chosen to reflect the hypothesis put forth in the third test of a shift in discourse for the unsplit form. Each extract exhibits how the split form continues in the same direction and the unsplit creates a shift. Were this test to be extended in detail to more works, I am confident that a similar trend would arise. Nevertheless, a short example from Exodus 12:22 shows how the split and unsplit appear together and function as laid out in the hypothesis.
Vang het bloed van het lam op in een schaal, doop daar een bosje hyssop in en strijk daarmee het bloed aan de posten van jullie huisdeuren. Daarna mag niemand het huis verlaten tot de morgen aanbreekt.


Take a bunch of hyssop, dip it into the blood in the basin and put some of the blood on the top and on both sides of the doorframe. None of you shall go out of the door of your house until morning.

(https://www.biblegateway.com/passage/?search=Exodus%2012&version=NIV)

The first split pronominal adverb is the continuation of the instructions of what to do in order to create the necessary marking to protect the people of Israel from the wrath to come. The unsplit form appears shortly afterward instructing them what to do with this blood. The preparation of the blood was the first step, yet without putting the blood on the doorframes, those in the houses would be not spared. Again, the split form continues the discourse (in this case the step by step instructions on how to prepare the blood) and then the unsplit shows the shift: use this for something that will save your lives.

In conclusion, I would like to emphasize that Yishai Tobin's concept of semantic integrality may be viewed as supporting the hypothesis set forth in this dissertation. Not only has semantic integrality shown that the perceived connectivity between clauses with the split form displays a continuous spatio-temporal-existential cline, but it also highlights how discourse continues. With unsplit forms, unmarked for semantic integrality, the reader/listener is provided with a signal that the information in the proceeding clause will most likely stand in contrast to
what has been occurring up to that point. This latter point makes itself clear through each extract from the two novels used. Further detailed work with other works of literature and corpus data would most likely continue to exhibit these tendencies for the two forms. Hermans provides a sound foundation, upon which to build this hypothesis, and were it possible to expand the breadth of the data compiled from 38-million-word corpus from the previous tests, I am sure that this qualitative analysis would fit there also. The two forms are chosen, possibly subconsciously, to indicate something, and minimizing this choice to simplistic parameters such as written vs. spoken does injustice not only to the authors but also to the magnificent creativity of language itself.
9 Conclusion

"Grammar is not semantically arbitrary. On the contrary, grammatical distinctions are motivated (in the synchronic sense) by semantic distinctions; every grammatical construction is a vehicle of a certain semantic structure; and this is its raison d’être, and the criterion determining its range of use" (Wierzbicka, 1988, p. 3).

The main goal of this dissertation has been to show that assumed synonymy between two similar grammatical forms can, and must be, questioned. Far too often grammars gloss over some of the more intricate differences in language, and for advanced users, these shortcomings lead the user/researcher to assume more than to understand. Hence, this dissertation has sought to confront these shortcomings and provide both a methodology and a collection of further reasons to continue questioning the assumed synonymy. As Bolinger states: "a difference in syntactic form always spells a difference in meaning" (1968, p. 127).

Crucial to this dissertation were the methods currently being employed in the field of Corpus Linguistics, for without the data, one can only speculate. For the first two tests, I relied heavily on quantitative data, reflecting the growing trend in linguistics to take more empirical approaches. Over time, it has become more important to base one's linguistic hypotheses on quantitative data and not assumptions and intuition. We have more access to real data, both spoken and written (whereby there is much more of the latter), so not to use it would seem counter-intuitive. As such, linguistics needs to begin approaching statistics with a more critical eye. This has been the case in sociolinguistics, and there is no reason other fields of linguistics should not follow suit. We need to be not only critical toward statistics, but our relationship as linguists to the field of statistics. There needs to be much more emphasis on teaching future linguists how to work with statistics and computer programming, should we desire to maintain
growth and development. Nevertheless, we must also remain connected to the human aspect of language, which is what makes the final test of this dissertation so important. For all of the statistical justifications for how one form appears over the other, there is still the underlying fact that a choice is made by a human. In this final section, I will review the previous tests, discuss some future research prospects, as well as further attempt to show that the data provided supports the hypothesis suggested.

Therefore, I agree with Bolinger (1968) and Goldberg (1995) here that different forms signal different meaning and use. Reviewing test one, I discovered that an earlier hypothesis on the cause of splitting cannot be entirely defended. Whereas van der Horst (1992) is correct in believing there to be a motivating factor for splitting, he lacks quantitative proof. One could argue that this lack of proof should not be seen as a fault of van der Horst, but rather the state of corpus technology at the time. While it may be true that one could accept that and move on, we are now able to dig deeper. With the proliferation of Big Data, it would be wise to use these various resources discover new problems and address older ones. The results shown in test one reflect how the verb itself cannot be considered the cause for splitting; something else must be motivating the split.

In test two, I show how looking deeper into the data can provide better insights into motivation. What once appeared as intuition has shown itself to be validated by raw data. Instead of looking at the construction itself, it became apparent that I needed to look further. One of the underlying premises of corpus linguistics is that frequency and certain statistical tests, when utilized properly, can show how and possibly why we use language as we do. Therefore, test two focused on frequency and attraction of each verb found in the constructions. Not only were these verbs individually analyzed based on their meanings, but also on their semantic categorization,
as provided in the Frequency Dictionary of Dutch. Test two shows that there is basically no correlation between the split form and the frequency of the verb. Whether or not the verb in the clause appears more or less frequent plays no role in the split form. Yet, for the unsplit the data displays the opposite. A negative correlation materializes between higher frequency verbs and their likelihood to be used in an unsplit construction. I have also shown that verbs belonging to fiction and general genres occur more frequently with the unsplit form. This also suggests that verbs of lower frequency are more likely to appear with the unsplit form. Via test two I have found motivation for use: less frequently occurring verbs tend to collocate with the unsplit form. Nevertheless, this motivation does not provide us with any meaning difference.

The final statistics based quantitative test seeks to use data to find meaning. For the first two tests, I have shown that data allow the linguist to make connections previously based on speculation and intuition. With test three, I seek to use similar methodologies to find meaning. Whereas describing meaning can be subjective, it was important to find parameters for test three which could allow for a more objective analysis. The use of Princeton's WordNet database permitted this objectivity. Due to the labor-intensive nature of test three, sample sizes were reduced to 281 in the split and 336 in the unsplit.

Revisiting Construction Grammar, I now posit that the the two structures below, suggested in chapter two, can be assigned meaning, depending on their use:

A) \([DAAR] [Preposition] \) [x]

B) \([DAAR] [x] [Preposition] \)

For structure A, I suggest that pronominal adverbs that occur in their unsplit form are mapped onto a meaning of POSSIBLE SHIFT IN DISCOURSE. Structure B maps onto a meaning consistent with NO SHIFT IN DISCOURSE. This form-meaning pairing has been observed not
just in the two forms of daarmee, as suggested in test three, but also, as test four shows, in other forms of the split and unsplit pronominal adverbs with daar. What remains to be seen is how this can be further tested with hier and waar.

Another aspect of this dissertation that I believe would warrant further study includes creating a questionnaire, whereby I would inquire native speakers of Dutch, in both the Netherlands and Belgium, as to what form they would expect in a given situation. For each question, I would seek narrative and spoken passages from novels and provide both a split and unsplit variant of the passage. Passages used would often include crucial scenes in stories with enough context to let the reader know that something drastic is to transpire. From here, the questionnaire would ask whether they would expect one or the other. Whereas I firmly believe that this test would further support our theory, one should adhere to a wise maxim, "Ask not what your language can do for your theory, but what your theory can do for your language".
## Appendix A: Tree Tagger Numerical Values

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>verbal infinitive of an intransitive used verb</td>
</tr>
<tr>
<td>210</td>
<td>verbal infinitive of a transitive used verb</td>
</tr>
<tr>
<td>220</td>
<td>verbal infinitive of reflexive used verb</td>
</tr>
<tr>
<td>230</td>
<td>verbal infinitive of helping verb</td>
</tr>
<tr>
<td>206</td>
<td>undeclined past participle of an intransitive used verb</td>
</tr>
<tr>
<td>216</td>
<td>undeclined past participle of a transitive used verb</td>
</tr>
<tr>
<td>226</td>
<td>undeclined past participle of reflexive used verb</td>
</tr>
<tr>
<td>236</td>
<td>undeclined past participle of helping verb</td>
</tr>
<tr>
<td>242</td>
<td>2 person form of an intransitive used verb</td>
</tr>
<tr>
<td>243</td>
<td>3 person form of an intransitive used verb</td>
</tr>
<tr>
<td>244</td>
<td>plural present of an intransitive used verb</td>
</tr>
<tr>
<td>245</td>
<td>singular preterite of an intransitive used verb</td>
</tr>
<tr>
<td>246</td>
<td>plural preterite of an intransitive used verb</td>
</tr>
<tr>
<td>247</td>
<td>imperative w/o personal pronoun of an intransitive used verb</td>
</tr>
<tr>
<td>248</td>
<td>imperative w/ personal pronoun of an intransitive used verb</td>
</tr>
<tr>
<td>251</td>
<td>1 person form of a transitive used verb</td>
</tr>
<tr>
<td>252</td>
<td>2 person form of a transitive used verb</td>
</tr>
<tr>
<td>253</td>
<td>3 person form of a transitive used verb</td>
</tr>
<tr>
<td>254</td>
<td>plural present of a transitive used verb</td>
</tr>
<tr>
<td>255</td>
<td>singular preterite of a transitive used verb</td>
</tr>
<tr>
<td>256</td>
<td>plural preterite of a transitive used verb</td>
</tr>
<tr>
<td>257</td>
<td>imperative w/o personal pronoun of a transitive used verb</td>
</tr>
<tr>
<td></td>
<td>Description</td>
</tr>
<tr>
<td>---</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>258</td>
<td>imperative w/ personal pronouns of a transitive used verb</td>
</tr>
<tr>
<td>261</td>
<td>1 person form of a reflexive used verb</td>
</tr>
<tr>
<td>262</td>
<td>2 person form of a reflexive used verb</td>
</tr>
<tr>
<td>263</td>
<td>3 person form of a reflexive used verb</td>
</tr>
<tr>
<td>264</td>
<td>plural present of a reflexive used verb</td>
</tr>
<tr>
<td>265</td>
<td>singular preterite of a reflexive used verb</td>
</tr>
<tr>
<td>266</td>
<td>plural preterite of a reflexive used verb</td>
</tr>
<tr>
<td>267</td>
<td>imperative w/o personal pronoun of a reflexive used verb</td>
</tr>
<tr>
<td>268</td>
<td>imperative w/ personal pronouns of a reflexive used verb</td>
</tr>
<tr>
<td>271</td>
<td>1 person form of a helping verb</td>
</tr>
<tr>
<td>272</td>
<td>2 person form of a helping verb</td>
</tr>
<tr>
<td>273</td>
<td>3 person form of a helping verb</td>
</tr>
<tr>
<td>274</td>
<td>plural present of a helping verb</td>
</tr>
<tr>
<td>275</td>
<td>singular preterite of a helping verb</td>
</tr>
<tr>
<td>276</td>
<td>plural preterite of a helping verb</td>
</tr>
<tr>
<td>277</td>
<td>imperative w/o personal pronoun of a helping verb</td>
</tr>
<tr>
<td>278</td>
<td>imperative w/ personal pronouns of a helping verb</td>
</tr>
</tbody>
</table>
### 11 Appendix B: Met Verbs

<table>
<thead>
<tr>
<th>aangaan</th>
<th>bedenken</th>
<th>bestraffen</th>
<th>corresponderen</th>
</tr>
</thead>
<tbody>
<tr>
<td>aanknopen</td>
<td>bedoelen</td>
<td>bestrijden</td>
<td>debatteren</td>
</tr>
<tr>
<td>aankomen</td>
<td>bedrogen</td>
<td>betitelen</td>
<td>delen</td>
</tr>
<tr>
<td>aanleggen</td>
<td>begaan</td>
<td>bezuren</td>
<td>dienen</td>
</tr>
<tr>
<td>aanspreken</td>
<td>begiftigen</td>
<td>bijstaan</td>
<td>doodgooien</td>
</tr>
<tr>
<td>aanvangen</td>
<td>beginnen</td>
<td>boffen</td>
<td>doodslaan</td>
</tr>
<tr>
<td>anvullen</td>
<td>begunstigen</td>
<td>botsen</td>
<td>doorgaan</td>
</tr>
<tr>
<td>aarzelen</td>
<td>behelpen</td>
<td>breken</td>
<td>doornemen</td>
</tr>
<tr>
<td>achterlaten</td>
<td>bekleden</td>
<td>chanteren</td>
<td>doorspekken</td>
</tr>
<tr>
<td>achtervolgen</td>
<td>bekopen</td>
<td>collaboreren</td>
<td>doorverbinden</td>
</tr>
<tr>
<td>adverteren</td>
<td>bekorten</td>
<td>combineren</td>
<td>doteren</td>
</tr>
<tr>
<td>afdoen</td>
<td>bekrachtigen</td>
<td>communiceren</td>
<td>dralen</td>
</tr>
<tr>
<td>affiliëren</td>
<td>bekronen</td>
<td>complimentieren</td>
<td>driejen</td>
</tr>
<tr>
<td>afgeven</td>
<td>belasten</td>
<td>committeren</td>
<td>duelleren</td>
</tr>
<tr>
<td>aflopen</td>
<td>bellen</td>
<td>concurren</td>
<td>dwepe</td>
</tr>
<tr>
<td>afnemen</td>
<td>belonen</td>
<td>condoleren</td>
<td>eindigen</td>
</tr>
<tr>
<td>afrekenen</td>
<td>bemoeien</td>
<td>confereren</td>
<td>engageren</td>
</tr>
<tr>
<td>afschepen</td>
<td>beogen</td>
<td>confronteren</td>
<td>eren</td>
</tr>
<tr>
<td>afwisselen</td>
<td>beraadslagen</td>
<td>congrueren</td>
<td>falen</td>
</tr>
<tr>
<td>antwoorden</td>
<td>bereiden</td>
<td>contrasteren</td>
<td>farceren</td>
</tr>
<tr>
<td>associëren</td>
<td>besluiten</td>
<td>converseren</td>
<td>feliciteren</td>
</tr>
<tr>
<td>bedelen</td>
<td>bestaan</td>
<td>correleren</td>
<td>flirten</td>
</tr>
<tr>
<td>gaan</td>
<td>inlaten</td>
<td>leuren</td>
<td>mikken</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>--------</td>
<td>----------</td>
</tr>
<tr>
<td>gebeuren</td>
<td>instemmen</td>
<td>leven</td>
<td>minnekozen</td>
</tr>
<tr>
<td>gelegen</td>
<td>integreren</td>
<td>lichten</td>
<td>mixen</td>
</tr>
<tr>
<td>gelijkloopen</td>
<td>inzitten</td>
<td>lijmen</td>
<td>morrelen</td>
</tr>
<tr>
<td>gelijkmaken</td>
<td>kampen</td>
<td>lokken</td>
<td>morsen</td>
</tr>
<tr>
<td>gelijkschakelen</td>
<td>kappen</td>
<td>lopen</td>
<td>naaien</td>
</tr>
<tr>
<td>gelijkstaan</td>
<td>kennismaken</td>
<td>maken</td>
<td>nokken</td>
</tr>
<tr>
<td>gelijkstellen</td>
<td>kibbelen</td>
<td>malen</td>
<td>occuperen</td>
</tr>
<tr>
<td>gelukwensen</td>
<td>kijven</td>
<td>manipuleren</td>
<td>omgaan</td>
</tr>
<tr>
<td>geuren</td>
<td>klaarkomen</td>
<td>marchanderen</td>
<td>omgorden</td>
</tr>
<tr>
<td>goedmaken</td>
<td>klaarstaan</td>
<td>markeren</td>
<td>omkopen</td>
</tr>
<tr>
<td>groeien</td>
<td>kletsen</td>
<td>matigen</td>
<td>omringen</td>
</tr>
<tr>
<td>haasten</td>
<td>klikken</td>
<td>meedoen</td>
<td>omspringen</td>
</tr>
<tr>
<td>handelen</td>
<td>klinken</td>
<td>meegaan</td>
<td>onderhouden</td>
</tr>
<tr>
<td>harmonieren</td>
<td>knoeien</td>
<td>meekomen</td>
<td>ontbijten</td>
</tr>
<tr>
<td>harrewarren</td>
<td>koketteren</td>
<td>meelevens</td>
<td>ophebben</td>
</tr>
<tr>
<td>hechten</td>
<td>korten</td>
<td>meelopen</td>
<td>ophouden</td>
</tr>
<tr>
<td>helpen</td>
<td>kortsluiten</td>
<td>meepraten</td>
<td>opknappen</td>
</tr>
<tr>
<td>heulen</td>
<td>kruisen</td>
<td>meevloeren</td>
<td>oplopen</td>
</tr>
<tr>
<td>huwen</td>
<td>lachen</td>
<td>meewerken</td>
<td>oplusteren</td>
</tr>
<tr>
<td>identificeren</td>
<td>laven</td>
<td>menen</td>
<td>opschepen</td>
</tr>
<tr>
<td>inenten</td>
<td>legeren</td>
<td>mengen</td>
<td>opschieten</td>
</tr>
<tr>
<td>infecteren</td>
<td>legitimeren</td>
<td>meten</td>
<td>optrekken</td>
</tr>
<tr>
<td>opzadelen</td>
<td>redeneren</td>
<td>slagen</td>
<td>telefoneren</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>overeenbrengen</td>
<td>redetwisten</td>
<td>slapen</td>
<td>terugbrengen</td>
</tr>
<tr>
<td>overeenkomen</td>
<td>rekenen</td>
<td>slijmen</td>
<td>terugvallen</td>
</tr>
<tr>
<td>overeenstemmen</td>
<td>rijden</td>
<td>sluiten</td>
<td>tevredenstellen</td>
</tr>
<tr>
<td>overladen</td>
<td>rijmen</td>
<td>smoezen</td>
<td>tobben</td>
</tr>
<tr>
<td>overleggen</td>
<td>rivaliseren</td>
<td>spelen</td>
<td>toekomen</td>
</tr>
<tr>
<td>overschrijden</td>
<td>roddelen</td>
<td>sporen</td>
<td>toerusten</td>
</tr>
<tr>
<td>overspoelen</td>
<td>rommelen</td>
<td>spotten</td>
<td>tooien</td>
</tr>
<tr>
<td>overstromen</td>
<td>rondkomen</td>
<td>spreken</td>
<td>treuzelen</td>
</tr>
<tr>
<td>overvallen</td>
<td>ruilen</td>
<td>staan</td>
<td>troosten</td>
</tr>
<tr>
<td>pacteren</td>
<td>samengaan</td>
<td>staven</td>
<td>trouwen</td>
</tr>
<tr>
<td>paraderen</td>
<td>samenhangen</td>
<td>stemmen</td>
<td>twisten</td>
</tr>
<tr>
<td>paren</td>
<td>samensmelten</td>
<td>steunen</td>
<td>uitbreiden</td>
</tr>
<tr>
<td>plooien</td>
<td>samenspannen</td>
<td>stijgen</td>
<td>uitdossen</td>
</tr>
<tr>
<td>plukharen</td>
<td>samenvallen</td>
<td>stoeien</td>
<td>uithalen</td>
</tr>
<tr>
<td>polemiseren</td>
<td>samenwerken</td>
<td>stofferen</td>
<td>uitkijken</td>
</tr>
<tr>
<td>praten</td>
<td>samenzweren</td>
<td>stoppen</td>
<td>uitkomen</td>
</tr>
<tr>
<td>profileren</td>
<td>scharrelen</td>
<td>straffen</td>
<td>uitmaken</td>
</tr>
<tr>
<td>pronken</td>
<td>schermen</td>
<td>strijden</td>
<td>uitmonsteren</td>
</tr>
<tr>
<td>puzzelen</td>
<td>schieten</td>
<td>strijken</td>
<td>uitpakken</td>
</tr>
<tr>
<td>rammelen</td>
<td>schikken</td>
<td>stroken</td>
<td>uitpraten</td>
</tr>
<tr>
<td>reageren</td>
<td>sieren</td>
<td>sympathiseren</td>
<td>uitrusten</td>
</tr>
<tr>
<td>redden</td>
<td>sjacheren</td>
<td>talmen</td>
<td>uitscheiden</td>
</tr>
</tbody>
</table>
uitvechten uitwisselen vaccineren vastleggen vechten venten verbazen verbinden verblijden verbluffen verbroederen verdienen verdragen verdrijven vereenzelvigen vereffenen verenigen vereren vergalopperen vergelijk
verschillen verschonen versieren versmelen warmen versnijden wedijveren verstaan weglopen vervolgen werken vervullen winnen verwarren wisselen verwelkomen woekeren verwisselen worstelen verzoenen zakken voeden zegevieren voeren zeulden volstaan zoekbrengen voorhebben zwaaien voortmaken vorderen
### Appendix C: Prepositions from Zwarts (1997)

<table>
<thead>
<tr>
<th>A Group</th>
<th>Prepositions that can create pronominal adverbs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dutch</strong></td>
<td><strong>English</strong></td>
</tr>
<tr>
<td>aan</td>
<td>at</td>
</tr>
<tr>
<td>achter</td>
<td>behind, after, beyond</td>
</tr>
<tr>
<td>bij</td>
<td>by, near, with</td>
</tr>
<tr>
<td>binnen</td>
<td>within, inside</td>
</tr>
<tr>
<td>boven</td>
<td>above, over</td>
</tr>
<tr>
<td>buiten</td>
<td>outside</td>
</tr>
<tr>
<td>door</td>
<td>through, by</td>
</tr>
<tr>
<td>in</td>
<td>in, inside, into</td>
</tr>
<tr>
<td>langs</td>
<td>along, past</td>
</tr>
<tr>
<td>met</td>
<td>with</td>
</tr>
<tr>
<td>na</td>
<td>after</td>
</tr>
<tr>
<td>naar</td>
<td>to, toward, for</td>
</tr>
<tr>
<td>naast</td>
<td>next to, beside, alongside</td>
</tr>
<tr>
<td>om</td>
<td>around, about</td>
</tr>
<tr>
<td>onder</td>
<td>under, underneath, beneath</td>
</tr>
<tr>
<td>op</td>
<td>on, upon, at</td>
</tr>
<tr>
<td>over</td>
<td>over, across, about</td>
</tr>
<tr>
<td>tegen</td>
<td>against</td>
</tr>
<tr>
<td>tegenover</td>
<td>opposite</td>
</tr>
<tr>
<td>tot</td>
<td>till, until, as far as, up to</td>
</tr>
<tr>
<td>tussen</td>
<td>between, among, amid</td>
</tr>
<tr>
<td>uit</td>
<td>out of, from</td>
</tr>
<tr>
<td>van</td>
<td>of, from</td>
</tr>
<tr>
<td>voor</td>
<td>before, in front of, for</td>
</tr>
<tr>
<td>voorbij</td>
<td>past, beyond</td>
</tr>
<tr>
<td>B Group</td>
<td>Prepositions that cannot create pronominal adverbs</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>Dutch</td>
<td>English</td>
</tr>
<tr>
<td>à</td>
<td>a, to</td>
</tr>
<tr>
<td>aangaande</td>
<td>concerning</td>
</tr>
<tr>
<td>behoudens</td>
<td>barring</td>
</tr>
<tr>
<td>beneden</td>
<td>under, below, underneath, beneath</td>
</tr>
<tr>
<td>benevens</td>
<td>in addition to</td>
</tr>
<tr>
<td>benoorden</td>
<td>north of</td>
</tr>
<tr>
<td>beoosten</td>
<td>east of</td>
</tr>
<tr>
<td>betreffende</td>
<td>regarding</td>
</tr>
<tr>
<td>bewesten</td>
<td>west of</td>
</tr>
<tr>
<td>bezijden</td>
<td>south of</td>
</tr>
<tr>
<td>bezuiden</td>
<td>as appears from</td>
</tr>
<tr>
<td>blijkens</td>
<td>in conformity with</td>
</tr>
<tr>
<td>conform</td>
<td>contra</td>
</tr>
<tr>
<td>contra</td>
<td>thanks to</td>
</tr>
<tr>
<td>dankzij</td>
<td>excluding</td>
</tr>
<tr>
<td>exclusief</td>
<td>during</td>
</tr>
<tr>
<td>gedurende</td>
<td>given</td>
</tr>
<tr>
<td>gegeven</td>
<td>as witness</td>
</tr>
<tr>
<td>getuige</td>
<td>considering</td>
</tr>
<tr>
<td>halverwege</td>
<td>halfway through, down, across</td>
</tr>
<tr>
<td>hangende</td>
<td>pending</td>
</tr>
<tr>
<td>inclusief</td>
<td>including</td>
</tr>
<tr>
<td>ingevolge</td>
<td>in pursuance of</td>
</tr>
<tr>
<td>inzake</td>
<td>on the subject of</td>
</tr>
<tr>
<td>jegens</td>
<td>toward</td>
</tr>
<tr>
<td>krachtens</td>
<td>on the strength of</td>
</tr>
<tr>
<td>luidens</td>
<td>according to</td>
</tr>
<tr>
<td>middels</td>
<td>by means of</td>
</tr>
<tr>
<td>naargelang</td>
<td>in proportion to</td>
</tr>
<tr>
<td>nabij</td>
<td>near</td>
</tr>
<tr>
<td>namens</td>
<td>in the name of</td>
</tr>
<tr>
<td>niettegenstaande</td>
<td>notwithstanding</td>
</tr>
<tr>
<td>nopens</td>
<td>concerning</td>
</tr>
<tr>
<td>omstreeks</td>
<td>in the neighborhood of</td>
</tr>
<tr>
<td>omtrent</td>
<td>with regard to</td>
</tr>
<tr>
<td>Dutch</td>
<td>English</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>ondanks</td>
<td>despite</td>
</tr>
<tr>
<td>ongeacht</td>
<td>regardless</td>
</tr>
<tr>
<td>onverminderd</td>
<td>without prejudice to</td>
</tr>
<tr>
<td>overeenkomstig</td>
<td>in correspondence with</td>
</tr>
<tr>
<td>per</td>
<td>per, by</td>
</tr>
<tr>
<td>qua</td>
<td>as, as far as</td>
</tr>
<tr>
<td>richting</td>
<td>in the direction of</td>
</tr>
<tr>
<td>rond</td>
<td>round</td>
</tr>
<tr>
<td>sedert</td>
<td>since</td>
</tr>
<tr>
<td>sinds</td>
<td>since</td>
</tr>
<tr>
<td>staande</td>
<td>pending</td>
</tr>
<tr>
<td>te</td>
<td>at, in</td>
</tr>
<tr>
<td>tijdens</td>
<td>during</td>
</tr>
<tr>
<td>trots</td>
<td>in defiance of</td>
</tr>
<tr>
<td>uitgezonderd</td>
<td>expected</td>
</tr>
<tr>
<td>vanwege</td>
<td>because of</td>
</tr>
<tr>
<td>versus</td>
<td>versus</td>
</tr>
<tr>
<td>via</td>
<td>via</td>
</tr>
<tr>
<td>volgens</td>
<td>according to</td>
</tr>
<tr>
<td>wegen</td>
<td>on account of, because of</td>
</tr>
<tr>
<td>zonder</td>
<td>without</td>
</tr>
</tbody>
</table>
13 Appendix D: Sources of the 38 Million Word Corpus

Alternative medicine and lifestyle


**Linguistics**


- Monthly journal from the society of "Our Language". 60th up to 64th year, 5 issues a year.

**Popular Science**

*Zenit*. Utrecht, 1992 nr. 6 en 1993 nr. 1 up to 7/8.

- Popular science monthly about astronomy, meteorology, space travel, space exploration and related sciences and techniques.

**Environment**


- Quarterly available to donors.

• From the national organization for the conservation of the Wadden Sea.


• Compiled by the Meander foundation

_Politics_


_Rooie vrouwen Magazine_. Rooie Vrouwen in de Partij van de Arbeid. Amsterdam, 1992 nr. 3; 1993 nr. 3 up to 9; 1994 nr. 1 up to 3.

• Magazine for women in the Workers' Party


_Reports of the plenary assembly of the House of Representatives of the Netherlands._

_Stenographic service of the House of Representatives of the Netherlands._ Den Haag, November and December 1995.


_Social Sciences_


Sports


Water Sports


- 4 selections per year- from the Royal Netherlandic tourism confederation

Legal corpus (not used)

14 Bibliography


