

The semantics of P in prepositional predicates

An investigation using corpus, cross-linguistic, and diachronic data

by

Iris Faber
s4509439

MA degree programme in
Linguistics and Communication Sciences (research)

31 May 2023, Nijmegen

Supervisors:
Dr. Peter de Swart
Dr. Margit Rem

Radboud University



Table of Contents

| | |
|---|----|
| Abstract..... | 1 |
| Chapter 1. Prepositional predicates: an overview..... | 2 |
| 1.1 What are prepositional predicates? | 2 |
| 1.2 Identifying prepositional predicates..... | 3 |
| 1.3 The semantics of P in previous literature..... | 4 |
| 1.3.1 The semantics of P outside of verb-preposition combinations | 5 |
| 1.3.2 Meaningless prepositions in prepositional predicates | 5 |
| 1.3.3 Contradicting evidence/literature..... | 7 |
| 1.4 The aim and structure of this thesis | 10 |
| Chapter 2. Predicate Selection | 12 |
| 2.1 ANS list..... | 12 |
| 2.2 Selection criteria | 13 |
| 2.2.1 Multi-word predicates..... | 13 |
| 2.2.2 Separable phrasal verbs | 13 |
| 2.2.3 Modern concepts..... | 13 |
| 2.2.4 Annotation summary..... | 14 |
| 2.3 Predicate frequency..... | 15 |
| 2.4 Loan words | 16 |
| Chapter 3. Semantic similarity of Dutch corpus data | 18 |
| 3.1. Introduction..... | 18 |
| 3.2. Data..... | 19 |
| 3.3. Analysis..... | 20 |
| 3.4. Discussion..... | 25 |
| Chapter 4. Cross-linguistic comparison..... | 26 |
| 4.1. Introduction..... | 26 |
| 4.2. Data..... | 27 |
| 4.2.1 Translations..... | 27 |
| 4.2.2 Dissimilarities | 28 |
| 4.3. Analysis..... | 30 |
| 4.3.1 Multidimensional Scaling | 30 |
| 4.3.1.1 MDS: method..... | 31 |
| 4.3.1.2 MDS: Results..... | 31 |
| 4.3.2 Colour coded analysis of full dissimilarity matrix..... | 32 |
| 4.3.3 Translated prepositions | 35 |
| 4.4. Discussion..... | 37 |

| | |
|--|----|
| Chapter 5. Diachronic analysis of prepositional predicates..... | 40 |
| 5.1. Introduction..... | 40 |
| 5.2. The history of Dutch case marking and prepositions (Van der Horst, 2008)..... | 40 |
| 5.3. Method..... | 43 |
| 5.3.1 Verb selection..... | 43 |
| 5.3.2 Data..... | 44 |
| 5.3.2 Analysis..... | 45 |
| 5.4. Results..... | 45 |
| 5.4.1 Azen op ‘have one’s eye on’..... | 45 |
| 5.4.2 Lijken op ‘resemble’..... | 46 |
| 5.4.3 Verhalen op ‘recover from’..... | 47 |
| 5.4.4 Vertrouwen op ‘trust/rely on’..... | 48 |
| 5.4.5 Beschermen tegen ‘protect against’..... | 49 |
| 5.4.6 Kunnen tegen ‘can stand’..... | 50 |
| 5.4.7 Verwerpen tegen ‘defend against’..... | 50 |
| 5.4.8 Zich verzetten tegen ‘resist’..... | 51 |
| 5.4.9 Dromen van ‘dream of’..... | 51 |
| 5.4.10 Overtuigen van ‘convince of’..... | 52 |
| 5.4.11 Schrikken van ‘get scared by’..... | 52 |
| 5.4.12 Voorzien van ‘provide with’..... | 53 |
| 5.5. Discussion..... | 53 |
| Chapter 6. Conclusion..... | 56 |
| References..... | 57 |
| Appendices..... | 62 |
| Appendix A: List of selected predicates and frequencies..... | 62 |
| Appendix B: Frequencies of predicates for LSA..... | 65 |
| Appendix C: Keywords based on LSA..... | 68 |
| Appendix D: Translations of all predicates..... | 73 |
| Appendix E: Intra preposition group dissimilarity score..... | 77 |

Abstract

Literature on prepositional predicates often describes how the preposition in such constructions has undergone semantic bleaching—it does not express any spatial or temporal information, but essentially functions as a grammatical marker of the object of a verb. The choice of preposition is thought to be arbitrary and idiomatic. However, while the latter seems to be true—one cannot substitute a preposition for another, and the meaning of the predicate is not compositional, but lexically determined—this thesis questions whether the choice of preposition is truly arbitrary.

In three studies, the relationship between preposition and verb was explored, aiming to shed more light on the semantic value of the preposition in such constructions. First, modern corpus data were analysed in order to examine the semantic context in which predicates are found, using Latent Semantic Analysis. The results of this study were inconclusive. A more abstract analysis was hypothesised to likely yield more information in further studies. Second, Dutch predicates were compared to their English and German counterparts, revealing that predicates group together similarly across language, disproving the idea that the choice of preposition is arbitrary. Furthermore, the translational equivalent of the Dutch preposition used was in all cases the most frequently used preposition in English and German (e.g., *tegen*, *against*, and *gegen*, respectively). Finally, diachronic data was used to investigate the evolution of a smaller subset of predicates. The data showed that that was much variation in how the predicate developed over time, but also showed some regularities within preposition groups.

To conclude, claims that the preposition is truly meaningless are proven untrue, as are statements on the idiosyncratic nature of the choice of a preposition. While most prepositions in prepositional predicates don't carry their original spatial or temporal meaning, they *are* selected based on the semantics of both the verb and (a metaphorical interpretation of) that preposition.

Chapter 1. Prepositional predicates: an overview

1.1 What are prepositional predicates?

Prepositional predicates are one of the more mysterious lexical categories, although they are well-represented in our vocabulary.¹ The sentences (1) to (3) give a few examples of Dutch prepositional predicates and their complements (with the predicates in bold and the complements underlined).

- (1) *Alex **wachtte op** de trein.*
“Alex was **waiting for** the train.”
- (2) *Elk product moet **voldoen aan** de kwaliteitsnormen.*
“Every product must **meet** the quality standards.”
- (3) *Je immuunsysteem **beschermt** je lichaam **tegen** ziektes.*
“Your immune system **protects** your body **against** disease.”

The most distinct (and defining) characteristic of a prepositional predicate is its preposition. As you can see, verbs can combine with a number of different prepositions to form such predicates (not limited to the ones exemplified here). These predicates can have just a single complement (the prepositional complement), as in (1) and (2), or multiple complements, as in (3). The preposition and verb form a single lexical unit, and the choice of preposition is lexically determined. In the past, studies have aimed to grasp a more thorough understanding of what a prepositional predicate entails—how it is formed, what the properties and characteristics are both of the verb and of the preposition, and how they have come to exist—yet many questions still remain to be answered.

Literature on prepositional complements often focuses on the question of how to determine whether a phrase is a prepositional complement or an adjunct, and on what separates the two. Van Heeswijck (1966) was among the first to ask this question in his paper aptly titled “Wat is het nu, voorwerp of bepaling?” [Which is it, object or adjunct?]. The *Codex van de spraakkundige terminologie* [Codex of Linguistic terminology] (1949) had changed the official designation to “voorzetselvoorwerp” [prepositional complement], but Van Heeswijck was not sure whether this new term properly described the clause, as he wondered whether it was even an object at all. Mainly through the semantic argument that prepositional complements can often be paraphrased to “normal” objects he arrives at the conclusion that prepositional complements are indeed objects.

¹ In this thesis I will make a deliberate distinction between the terms *verb* and *predicate*, in order to retain some semblance of order in the many times either term is used. The former term refers to a part of speech that needs no further explanation and will be used in its normal use. The term *predicate*, as described in the Concise Oxford Dictionary of Linguistics, is used in literature with two different meanings (Matthews, 2007). First, in traditional grammar it is generally used to refer to the part of the sentence representing “what is said, or *predicated of* the subject”, including any objects, predicatives, and/or adjuncts. The second notion defines a predicate as the main verb, together with any auxiliaries, certain prepositions, and other phrasal particles, but excluding arguments. I choose to adhere to this latter notion of the term *predicate*, as I require a term for the combination of a verb and preposition, but have no need to include an argument as well. Making this distinction provides me with separate terms to refer to the prepositional predicate as a whole, and to the verb that makes up part of it.

In the decades that followed, many other scholars have proposed and debated about different criteria to identify prepositional complements and to distinguish prepositional complements from adjuncts (amongst others, Broekhuis, 2004; Duinhoven, 1989; Paardekooper, 1986; Schermer-Vermeer, 1990; Van den Toorn, 1971; Van Driel, Van den Toorn, & Vullings, 1978; Zwaan, 1972). I will illustrate a number of such criteria presented in Broekhuis (2004) in section 1.2. Before I go into the specifics, however, it should be noted that these are not foolproof criteria. Rather, Broekhuis argues that no foolproof criteria *can* be formulated that apply to all and only prepositional complements, and therefore moves his attention to “necessary” and “sufficient” criteria. The former is a group of criteria that a prepositional phrase (PP) must meet in order to be called a prepositional complement, but that do not necessarily mean the phrase in question is one. The latter group are criteria that, if met, indicate that a PP is a prepositional complement, but not meeting them does not mean a PP is, in fact, not an prepositional complement.

1.2 Identifying prepositional predicates

Compare sentences (4a) and (4b). They look very similar—as can often be the case with complement PPs and adjunct PPs—but the former contains the prepositional predicate *wachten op* ‘wait for’ and its complement, while the latter contains the verb *wachten* ‘wait’ and an locational adjunct.

- | | | |
|-------|---|-----------|
| (4) a | <i>Jan wacht op zijn vader.</i> “John is waiting for his father.” | [object] |
| b | <i>Jan wacht op het perron.</i> “John is waiting on the platform.” | [adjunct] |
| | | |
| (5) a | <i>Jan wacht daar.</i> “John is waiting there.” | |
| b | <i>Jan wacht daarop.</i> “John is waiting for that.” | |

One of the criteria Broekhuis (2004) discusses is whether or not the prepositional phrase can be replaced with a locational adverb such as *hier* ‘here’ or *daar* ‘there’. Locational adjuncts can often be replaced, while prepositional complements cannot. Therefore, sentence (5a) can only correspond to (4b), not to (4a). This is not surprising to Broekhuis, as replacing the prepositional phrase in (4a) with *daar* causes the sentence to lose its preposition. On the other hand, if you replace the phrase with a pronominal adverb, like *daarop* in sentence (5b) (which is made up of the locational adverb *daar* ‘there’ and the preposition *op*) it can only correspond to the prepositional object in (4a), not to the adjunct in (4b). It then stands to reason to assume that it is the presence of a preposition that determines whether or not an adverb can replace a prepositional object.

Second, the preposition in a prepositional object phrase can, usually, not be replaced by another preposition. In contrast, the preposition in an locational adjunct can be replaced by others. Sentences (6a) and (6b) portray this. It should be noted that there are a number of verbs that can combine with more than one preposition, such as *jagen* ‘hunt’ with *op* and *naar*, but that often leads to a completely different interpretation of the verb: *jagen op* has a

literal interpretation ‘hunt for’ (e.g., ‘hunting for deer’), while *jagen naar* has the metaphorical interpretation ‘pursue’ (e.g., ‘pursuing success’).

- (6) a *Jan wacht op/*naast/*bij zijn vader.*
“John is waiting for/*next to/*at his father.”
- b *Jan wacht op/naast/bij het perron.*
“John is waiting on/next to/at the platform.”

Another clue that can point us in the direction of either argument or adjunct is word order. As prepositional predicates are part of the lexical projection of the verb, they are part of the VP. Prepositional adjuncts, on the other hand, are not. Therefore, in a neutral sentence with no added emphasis, prepositional complements are closer to a sentence final verb than prepositional objects in the same sentence. The examples in (7) illustrate this; the prepositional complement *op zijn vader* ‘for his father’ needs to be closer to the verb *wacht* ‘wait’ than the adverbial phrase *op het perron* ‘on the platform’, as in (7a), while the reverse order in (7b) is ungrammatical (unless you interpret *op zijn vader* as an additional locational adjunct, which, though not impossible, would be unlikely).

- (7) a ... *dat Jan [op het perron]_{adv} [op zijn vader]_{compl} wacht.*
... that John is waiting for his father on the platform.
- b *... *dat Jan [op zijn vader]_{compl} [op het perron]_{adv} wacht.*
... that John is waiting on the platform for his father.

Although these tests seem rather straightforward, there has been, and still is, much disagreement on the definition and identification of these phrases. In fact, the paper by Broekhuis (2004), which the criteria discussed above come from, sparked a passionate debate between the author and Schermer-Vermeer across several articles in *Nederlandse Taalkunde* (Broekhuis, 2004; Schermer-Vermeer, 2006; Broekhuis, 2007; Schermer-Vermeer, 2007), in which the predicate *worstelen met* ‘struggle with’ is investigated as a means of discussing the classification of prepositional predicates and identifying characteristics of prepositional complements.

Such discussions illustrate the debate around identifying prepositional predicates. There is, however, one criterium that more researchers have agreed upon: the idea that the preposition in a prepositional phrase is semantically bleached. In order to further discuss that idea, section 1.3.1 will first establish what semantic value prepositions ordinarily carry. Section 1.3.2 will then elaborate on the stance that the preposition in prepositional phrases has no meaning, before discussing contradicting viewpoints in 1.3.3.

1.3 The semantics of P in previous literature

The analyses of the semantics of prepositions in prepositional predicates can roughly be divided into two camps: those that say P is meaningless, and those that say it is not. In order to determine if and how the semantics of P in a prepositional predicate differ from other uses, we first need to establish what semantic value P can carry.

1.3.1 The semantics of *P* outside of verb-preposition combinations

We can distinguish between several different kinds of prepositions (Beliën, 2021): first, there is a substantial number of spatial prepositions, which generally specify the location of object A relative to object B, or in other words, the location of a given Figure specified in relation to a Ground.² For example, if the Figure is in an internal location with respect to the Ground, both English and Dutch would use the preposition *in* (e.g., *het cadeau is in de doos/the gift is in the box*); in a situation where the Figure is located externally to the Ground and vertically above it, one would use the English preposition *on* if the Figure and Ground were touching (e.g., *there is a book on the table*) or *above* if there were not (e.g., *a lamp is hanging from the ceiling above the table*). Similarly, you would use respectively *op* and *boven* in Dutch.

These locational prepositions constitute one type of spatial prepositions. The other type are directional prepositions: these indicate a change in location and denote the path a Figure travels with relation to the Ground. Take the sentence *the cat jumped into the box*, for example. The preposition *into* indicates that the Figure, *the cat*, travelled some path from a location external to the Ground, *the box*, ending at an internal location to the Ground.

A different type of preposition are temporal ones. These, as the name suggests, express the position of an event in time. Some prepositions can only be used to express a temporal relationship, such as *since*, *during*, and *until*—or *sinds*, *tijdens*, and *tot* in Dutch—, but some spatial prepositions can also be used for temporal expressions, such as *in*, *at*, and *from ... to* (resp. *in March*, *at 5pm*, *from Monday to Friday*).

The *Algemene Nederlandse Spraakkunst* [General Dutch Linguistics] also distinguishes a group they call ‘other’ (Beliën, 2021): prepositions do not have a spatial nor a temporal meaning. These are prepositions like *with(out)*, *except*, and *of*. They can express a number of different meanings, including accompaniment, means, cause, and purpose.

These different categories of prepositions are summarised in Table 1.1; English and Dutch examples are provided for each type.

Table 1.1

Examples of English and Dutch spatial, temporal, and other prepositions

| Type | Subtype | English examples | Dutch examples |
|----------|-------------|--|--|
| Spatial | Locational | <i>in, at, on, under, between</i> | <i>in, bij, op, onder, tussen</i> |
| | Directional | <i>from ([...] to), into, towards, through</i> | <i>van ([...] naar), in, richting, door ([...] heen)</i> |
| Temporal | | <i>since, during, until, after</i> | <i>sinds, tijdens, tot, na</i> |
| Other | | <i>with, without, except, of</i> | <i>met, zonder, behalve, van</i> |

1.3.2 Meaningless prepositions in prepositional predicates

Now, going back to the tests to distinguish between prepositional objects and adjuncts described in 1.2, many papers on the topic mention that the preposition in the latter construction retains its spatial/temporal/other meaning, while a preposition that functions as part of a prepositional predicate does not. You might have already noticed this in the example

² Other terms used for Figure and Ground are *Trajector* and *Landmark*, respectively, used in works like Lakoff (1987) and Langacker (1987).

sentences in this chapter. Sentence (8), for example, contains the preposition *op* ‘on’, but it does not have the locational meaning described in 1.3.1—you do not rely *on top of* someone physically. Instead, it indicates the person/object you can trust. Likewise, *naar* in (9) does not have the directional meaning of *towards*, but rather indicates the entity that is being searched for. Adversely, prepositional adjuncts with the same prepositions would express, respectively, locational and directional information.

- (8) *Je kan altijd **op mij vertrouwen**.*
 “You can always **rely on me**.”
- (9) *Julia **zoekt naar een oplossing** voor haar probleem.*
 “Julia **is looking for a solution** to her problem.”

It is not assumed that *naar* in (8) and *naar* in the directional sense are unrelated. Instead, the preposition is described to have grammaticalised. When a word (or phrase, morpheme, ...) is undergoing grammaticalisation, that often includes semantic change of that word—preserving its original function is optional (Hopper & Traugott 1993). Frequently, that semantic change comes in the form of semantic bleaching, or the reduction of the semantic content of a word. It seems that the preposition that accompanies these verbs has become largely grammaticalised and has lost semantic value. For this reason, Broekhuis, Corver and Vos (2015) refer to them as “functional prepositions” (p. 284).

However, to what degree these prepositions have grammaticalised exactly remains unspecified. Are they completely void of meaning and purely syntactic? Do they retain some semantic value, just not of the spatial or temporal type? Has their meaning evolved to a more metaphorical one?

A collection of quotes is listed below, showing different authors’ take on the meaningless preposition (emphasis added). While all assume or argue for a meaningless preposition, it does not seem like all agree on what that entails. “Semantically empty” (Somers, 2009) and “[the preposition] is really meaningless” (Ackema, 2015) are hardly the same as “the locational/temporal meaning is lost” (Broekhuis, 2004) or “a reduced semantic transparency of the preposition” (Vandeweghe, 2015), after all.

Somers (2009, p. 514): “such prepositions seem to be *semantically empty*”

Engel & Schumacher (1976, p. 62): “generally [the preposition] has *no meaning of its own*”

Heringer (1968, p. 434-435): “It is easy to establish however that the preposition in the prepositional object too has often *only syntactic significance*... Since the preposition does not enter into a paradigmatic opposition, *no semantic significance* can be ascribed to it.”

Ackema (2015, p. 266-267): “the preposition in these cases [when used as complements, not adjuncts] is *semantically bleached*: it does not express a location, time, direction, or anything like that anymore, but is really meaningless”

Broekhuis (2004, p. 97): “het voorzetsel [heeft] in veel gevallen *geen duidelijke semantische inhoud*” [the preposition has, in many cases, *no clear semantic value*]

Ibid. (p. 102): “[...] *de locationele/temporele betekenis* van het voorzetsel in isolatie *gaat verloren.*” [*the locational/temporal meaning* of the preposition in isolation *is lost.*]

Broekhuis, Corver & Vos (2015, p. 284): “Furthermore, the prepositions normally *do not have a welldefined semantic content* [...]”

Vandeweghe (2015, p. 456): “De ‘vaste’ verbinding impliceert *een verminderde semantische doorzichtigheid van het voorzetsel*, wat resulteert in verminderde vervangbaarheid, uitbreidbaarheid, beklemtoonbaarheid.” [The ‘fixed’ connection [between verb and preposition] implies a *reduced semantic transparency* of the preposition, which results in a reduced ability to be replaced, extended and stressed.]

Authors also speak on the randomness of the choice of preposition. Hoeksema (2014), for example, says: “Bij voorzetselvoorwerpen is de combinatie van voorzetsel en werkwoord *vaak arbitrair en idiomatisch*” [With prepositional predicates, the combination of preposition and verb is *often arbitrary and idiomatic*]. Van Bart & Sturm (1987, p. 169) go one step further and say: “[The choice of the preposition is], as far as linguistic knowledge is concerned, *a completely idiosyncratic phenomenon*, that is to say, a phenomenon for which no rules can be formulated, because *every verb seems to select its preposition in an arbitrary way.*”

Now, I will not argue that authors are wrong in saying that the preposition has a reduced semantic value, as that much is obvious. But to what extent that meaning is lost, and whether or not the choice for a certain preposition is truly random remains an open question.

1.3.3 Contradicting evidence/literature

Some researchers disagree with the viewpoint explained in the previous section, and have argued in favour of the idea that even though the preposition has lost its spatial/temporal meaning, it is not void of semantic value (see, for example, Vandeweghe, 2020, who argues that both verb and preposition carry meaning, though that of the preposition is more abstract than that of the verb).

First, it does not seem to be the case that all prepositional predicates are formed with meaningless prepositions. In a paper on differential object marking, De Swart (2014) identifies a group of Dutch verbs that use a preposition with inanimate, but not animate, objects, as (10) and (11) illustrate. This preposition clearly retains its locational meaning, as *in* signifies the location of the biting. This would, according to general agreement, imply that *het brood* is not a prepositional object. However, the characteristics of the phrase pattern with prepositional objects on all other objecthood tests—these include, for example, the replacement test with *daar(in)*, and word order in verb final clauses, as described in section 1.2 of this chapter. This study provides evidence for the fact that prepositions in prepositional predicates *can* still carry their original locational meaning.

(10) *De hond beet de man.*
“The dog **bit** the man.”

(11) *De hond beet in het brood.*
“The dog **bit** the bread.”

Some predicates even only gain their meaning when combined with the preposition. Compare, for example, *geven* ‘give’ and *geven om* ‘care about/for’. The emotional aspect of the predicate is gained by the combination with *om*. The same happens with the English verb *believe* in (12a) and (12b) (example taken from Zato, 2014); the former means that John believes what Bill is saying, while the latter more or less means that John trusts Bill. That means, Zato argues, that the preposition *in* contributes meaning to the sentence.

- (12)a John always **believes** Bill.
- b John always **believes in** Bill.

Botwinik-Rotem (2004, 2011) explains the difference between the two sentences in a different way: she argues that the preposition *is* void of meaning, and the structural differences are due to the different semantic roles of the objects. So-called ‘underspecified’ roles, like targets and goals, must be realised as PPs, as happens in (12b). Zato (2014) disagrees, however, because while that might explain the difference between (12a) and (12b), it does not explain meaning differences in verbs that allow preposition alternation. Section 1.2 already mentioned one example of such a predicate: *jagen op/naar* ‘hunt for/pursue’, whose meaning changes based on the preposition used. Such examples can also be found in other languages, including English (ibid, p. 60). The verb *belong*, for example, can combine with *to* (13a), *on* (13b) and *in* (13c). The first of these constructions denotes possession, the second a locative relationship, where the appropriate place for the plate is on the table, and the final expresses a part/whole relationship.

- (13)a The book **belongs to** Mary.
- b The plate **belongs on** the table.
- c This page **belongs in** that book.

If the preposition were truly meaningless, it would be odd that it could not be replaced by another ‘meaningless’ preposition. After all, both *of* in (14) and *on* in (15) would have no semantic value and merely function to introduce the object of the verb. Since they can both fulfil this function, why can one not be replaced by another? It might be that the combination of verb and preposition is lexically determined, but using so many different prepositions is not economical.

- (14) I am taking care of my sick mother.
- (15) Please concentrate on the issue at hand.
- (16) *I am taking care on my sick mother.
- (17) *Please concentrate of the issue at hand.

It is believed that prepositions developed out of a need to mark the role of different constituents when case marking became obsolete. In Dutch, this happened in the period of Middle Dutch. Before case marking disappeared, four cases were used: nominative, genitive,

dative, and accusative. If one needed to distinguish between genitive, dative and accusative objects, prepositions would be an ideal candidate. Accusative objects would be unmarked, and both genitive and dative objects would become prepositional predicates. But, as Zwarts (2014) notes, then only two prepositions would be needed. However, in modern Dutch, verbs can be combined with at least fifteen different prepositions (Beliën, 2021): *aan, achter, bij, in, met, naar, om, onder, op, over, tegen, tot, uit, van, and voor*. And Dutch is not alone in using more than the needed number of prepositions. English uses at least eight (*about, against, at, for, from, in, on, to, and with*), as does French (*à/au/aux, contre, dans, de/des/d', en, par, pour, and sur*), and Spanish uses five (*a, con, de, en, por*). In German, a language in which case marking is still productive, fourteen different prepositions are used to form prepositional predicates (*an, auf, aus, bei, für, gegen, mit, nach, über, um, unter, von, vor, zu*).

Finally, it has been noted that there appear to exist some regularities in which preposition a verb takes. Semantically related verbs appear to combine with the same preposition more often than could be expected if the choice were truly random. Several clusters of semantically similar verbs can be identified when looking at Dutch prepositional predicates. For example, there are a number of emotional support verbs like *vertrouwen* ‘trust’, *bouwen* ‘rely’ (lit. ‘build’) and *steunen* ‘support’ that all combine with *op*. Verbs that convey a transformation (e.g., *veranderen* ‘change’, *transformeren* ‘transform’, and *evolueren* ‘evolve’) combine with *tot* and/or *in* (Loonen, 2003; Vandeweghe, 2020). The content of communication is indicated with *over*, while the addressee is indicated with *tot* or *tegen* (in, for example, *spreken* ‘speak’ and *praten* ‘talk’). Many verbs involving emotions (e.g., *lachen* ‘laugh’, *huilen* ‘cry’, *rouwen* ‘mourn’, and *malen* ‘care’) use *om*.

Similarly, there is a selection of action verbs that carry a negative meaning, like *protesteren* ‘protest’, *strijden* ‘fight’, *vechten* ‘fight’, and *verzetten* ‘resist’, that all combine with *tegen*. This might have something to do with some inherent negativeness in the meaning of *tegen* (most often translated with ‘against’). For instance, in the predicate *zijn tegen* ‘be against’, not the verb, but only the preposition brings the ‘negativity’. You can see a similar thing happening with the verb *pleiten* ‘plead’, which, on its own, has a mostly neutral or positive interpretation, but combined with *tegen* carries a negative meaning. The reverse is also true for the predicates *zijn voor* ‘be in favour of’ and *pleiten voor* ‘plead for/argue in favour of’, where the preposition *voor* ‘for’ adds positive connotations.

Loonen (2003) lists, in section 4.3 of his dissertation, many more such semantically similar groups, across fifteen different prepositions. He argues that while a preposition does express a certain relationship between phrases, as much previous literature has claimed its function is, it also maintains a semantic value that colours that relationship. Examples from this list are his meanings listed for *onder*, which can, according to Loonen, indicate (movement to) a position in the middle of a group or (movement to) a position lower than another named entity.

These semantic clusters are also investigated in Höllein (2019), who explores the topic of prepositional complements and adverbs in German. He aims to find semantically similar groups of predicates—or, as he calls them, ‘semantic niches’. To this end, he created a list of all verbs that take a prepositional objects, based on lexicon data. He supplements this list with other predicates that resulted from a corpus search for productive verb-preposition combinations. Finally, all 2000 predicates, combining with 17 different prepositions, are grouped based on the semantics of the predicate. It does raise an important question for how to establish the semantic similarity of predicates. He does so mostly based on his own intuitions, but checks the clusters against groups described in previous literature and dictionary/lexicon/grammar and corpus data.

Höllein hypothesises that the verb, preposition, and case together act as carriers of the niche meaning, but the preposition indicates the specific niche. The collected data reveals two things: the prepositional objects in the data encode specific semantic roles, and the use of a certain prepositional object does indeed indicate a significant semantic niche. For example, the preposition case combination ‘*an* + accusative’ was determined to encode the semantic niche Höllein named ‘Adressatum’ (addressee). In (18), this construction signifies to whom ‘he’ is writing (viz. *his colleague*) and in (19) the construction identifies to whom the books are delivered (viz. *the customer*). When you combine *an* with a different case, however, you have a different construction and therefore a different niche. The combination ‘*an* + dative’ encodes the semantic niche ‘Affiziert’, the affected party. (20) and (21) show example sentences within this niche.

- (18) *Er schreibt an seinen Kollegen.*
 “He is writing to his colleague.”
- (19) *Er liefert die Bücher an den Kunden.*
 “He delivers the books to the customer.”
- (20) *Sie bastelt an ihrem Auto.*
 “She is working on her car.”
- (21) *Sie verletzt sich an dem rostigen Stacheldraht.*
 “She injured herself on the rusty barbed wire.”

Based on this data, Höllein concludes that the prepositions are not meaningless, but are indicators of niches. He does note that the synchronous nature of his investigation limits the extent of what it reveals of prepositional predicates, as language change is natural and he finds that the prepositional object is “extremely dynamic” (p. 290). Finally, he discusses how niches cannot be translated 1:1 from one language to another.

This final point is also stated by Hoeksema (2014), Ackema (2015) and Saint-Dizier (2006), among others: different languages do not use the same preposition for the translational equivalent of the same verb, or even use prepositional objects for the same verbs; it varies idiosyncratically from language to language (Ackema 2015, p. 267). For example, *wachten op* (lit. ‘wait on’) is translated to the English *wait for*. Where *over* is used for Dutch communication verbs, *about* is used for the topic of discussion with English verbs (*talk/write/argue/boast/... about*). *Dromen over* (lit. ‘dream over’) translates to *dream about* in English, *revêr de* in French (lit. ‘dream of’) and *soñar con* (lit. ‘dream with’) in Spanish and Portuguese (Saint-Dizier 2006, p. 2).

1.4 The aim and structure of this thesis

In this thesis, I will investigate some of the open questions surrounding prepositional predicates, in an effort to shed some more light on the construction in general, and in particular the characteristics of the preposition that complements a verb. While prepositions do not carry their spatial/temporal meaning, I aim to provide evidence for the hypothesis that the semantic value of a preposition is not irrelevant. For that purpose, I will focus most of my attention on Dutch predicates. What predicates specifically will be investigated is discussed in Chapter 2.

Höllein (2019) measured semantic similarity by dividing the predicates into groups based on introspective data, supplemented by dictionary and lexicon data. In this thesis, I will use Latent Semantic Analysis, a computational method, to find semantically similar clusters of predicates in order to investigate whether such a method can supplement the data presented by Höllein and Loonen (2003), and identify any existing patterns within a set of Dutch prepositional predicates. This study investigating the relationship between semantic similarity of verbs and preposition choice will be discussed in Chapter 3 of this thesis.

Chapter 4 will describe a cross-linguistic study to investigate the claims that object marking with prepositions varies idiosyncratically from language to language and that the selection of a specific preposition is arbitrary. The Dutch predicates from the first study will be compared to their German and English translational equivalents, to see whether predicates group together similarly in these different languages, and to explore how case marking is utilised with the German predicates.

Finally, the diachronic characteristics of prepositional predicates will be examined in Chapter 5. For a selection of predicates, literature and dictionary data will provide insight in how they developed from Old Dutch/(Early) Middle Dutch into the modern predicates used today. Whatever changes might have occurred in the development of the construction (if, for example, a verb changed the preposition it combines with), can help provide information on *why* they changed and the underlying motivations for choice of preposition.

Chapter 6 will conclude and discuss the overall findings. Additionally, appendices are added containing the predicates under investigation and data pertaining to the respective studies.

Chapter 2. Predicate Selection

This chapter will describe the selection process for the prepositional predicates investigated in this thesis.

2.1 ANS list

In order to select predicates for the studies in this thesis, I used the list of Dutch prepositional predicates from the *ANS* (Haeseryn, Romijn, Geerts, de Rooij, & van den Toorn, 1997, section 20-6-2). The total number of predicates in this list is 744, divided over fifteen groups, based on the preposition they govern. The preposition groups and the number of predicates in each group are presented in Table 2.1. I aimed to select ten predicates from each group. The groups of predicates that govern the prepositions *achter* ‘behind’, *bij* ‘with’, *onder* ‘under’, and *uit* ‘out’ contain too few predicates to select from and are therefore excluded. A final group was also removed from consideration, namely the predicates that combine with *met* ‘with’. The status of phrases with *met* in the adjunct-argument discussion is not agreed upon by linguists, with some arguing for analysing *met*-PP’s as adverbial phrases (conf. Broekhuis, 2004, 2014) and others defending the view that they are arguments (conf. Vandeweghe & Coleman, 2011). As this is still a topic very much under discussion, I chose not to include these predicates.

This left us with ten preposition groups (619 predicates), which, with ten predicates per group to be selected, will provide hundred predicates in total. In order to make this selection, several selection criteria were employed, verb frequencies were looked up, and translations were checked, as described in sections 2.2 to 2.4.

Table 2.1

All preposition groups from the ANS (Haeseryn et al., 1997), the number of predicates they contain and whether or not they are represented in the present thesis.

| Preposition | Number of predicates | Included in analysis |
|-------------|----------------------|----------------------|
| Aan | 84 | yes |
| Achter | 3 | no |
| Bij | 15 | no |
| In | 36 | yes |
| Met | 93 | no |
| Naar | 35 | yes |
| Om | 24 | yes |
| Onder | 3 | no |
| Op | 111 | yes |
| Over | 63 | yes |
| Tegen | 42 | yes |
| Tot | 41 | yes |
| Uit | 11 | no |
| Van | 110 | yes |
| Voor | 73 | yes |

2.2 Selection criteria

2.2.1 Multi-word predicates

Multi-word predicates were excluded from the selection. ‘Multi-word’ was defined as a predicate consisting of more than two words—preposition included. Examples of multi-word predicates are *trouw zijn aan* ‘be loyal to’ *berouw hebben over* ‘repent’, and *onder de indruk zijn van* ‘be impressed with’. These phrases are difficult to search for in corpora and dictionary, and are thus excluded from selection.

One exception to this are predicates that consist of a reflexive verb, a reflexive pronoun, and a preposition (e.g. *zich onderwerpen aan* ‘submit to’). While those do contain more than two words, you do not need the reflexive pronoun to search for the predicate in a corpus, and they are therefore not problematic for the present studies.

2.2.2 Separable phrasal verbs

The second selection criterion is whether or not the predicates (a) have an identical counterpart that functions as a separable phrasal verb, and/or (b) are made up of a separable phrasal verb (and a preposition).

Some Dutch prepositional predicates could also be a separable phrasal verb. For example, the phrase *drinkt op* could be used both in the sentence *hij drinkt op de overwinning* ‘he drinks/toasts to the victory’ and the sentence *hij drinkt de thee op* ‘he drinks the tea (he finishes it)’. While those are different lemma’s—*drinken (op)* and *opdrinken*, respectively—, no corpus I could find that would be suitable for the following studies distinguishes between the two. Any search for this type of predicate would thus retrieve a possibly very large number of sentences that contain a verb different from the one intended.

The second option is that the verb itself is made up of a separable phrasal verb, such as *opkijken* in *opkijken naar* ‘look up to’. In a main clause, this would be used as *de jongen kijkt op naar zijn vader* ‘the boy looks up to his father’. As most automatic parsers do not recognise this as a separated phrasal verb, any corpus using such parsers would fail to identify a significant amount of sentences containing this verb, possible skewing the data. The only way to include these sentences would then be to include all sentences with *kijkt* ‘looks’, but that only leads to more difficulties (as *kijken* ‘look’ is a verb in and of itself and also other prepositional predicates, such as *kijken naar* ‘look at’, would be included).

Both types of predicates would produce unreliable results, and are therefore excluded from selection.

2.2.3 Modern concepts

Since part of this thesis will focus on historical language data, no predicates can be included that clearly relate to modern concepts. Fortunately, there were no such predicates in this dataset.

2.2.4 Annotation summary

Table 2.2 presents an overview of the annotations. For each group, the number of predicates that are annotated as having a certain feature are recorded. The features included are those described in sections 2.2.1 to 2.2.3. The feature *modern concept* is not included in the table, as no predicates were found belonging to this category. The criterion *separable phrasal verb* is first listed as a whole, and subsequently split among the two types mentioned in section 2.2.2. These latter columns are marked by darker shading. The final column of the table notes the number of predicates that are unproblematic, which is here defined as not having a positive annotation in either of the first two categories: *multi-word predicate* and/or *separable phrasal verb*. After removing all predicates that were problematic, all groups still included a large enough number of predicates to select from during the final round.

Table 2.2

An overview of the amount and percentage of verbs annotated with each feature, both per predicate group and for all predicates in total.

| Preposition | Total number of predicates | Multi-word predicate | Separable phrasal verb (total) | V+P also separable phrasal verb | Verb = separable phrasal verb | Unproblematic |
|--------------|----------------------------|----------------------|--------------------------------|---------------------------------|-------------------------------|---------------|
| <i>Aan</i> | 84 | 31 (36.90%) | 17 (20.24%) | 3 (3.57%) | 14 (16.67%) | 37 (44.05%) |
| <i>In</i> | 36 | 17 (47.22%) | 3 (8.33%) | 1 (2.78%) | 2 (5.56%) | 16 (44.44%) |
| <i>Naar</i> | 35 | 6 (17.14%) | 3 (8.57%) | 0 (0%) | 3 (8.57%) | 26 (74.29%) |
| <i>Om</i> | 24 | 8 (33.33%) | 0 (0%) | 0 (0%) | 0 (0%) | 16 (66.67%) |
| <i>Op</i> | 111 | 39 (35.14%) | 30 (27.03%) | 8 (7.21%) | 22 (19.82%) | 44 (39.64%) |
| <i>Over</i> | 63 | 21 (33.33%) | 8 (12.69%) | 1 (1.59%) | 7 (11.11%) | 34 (53.97%) |
| <i>Tegen</i> | 42 | 10 (23.81%) | 12 (28.57%) | 0 (0%) | 12 (28.57%) | 20 (47.62%) |
| <i>Tot</i> | 41 | 10 (24.39%) | 8 (19.51%) | 0 (0%) | 8 (19.51%) | 23 (56.10%) |
| <i>Van</i> | 110 | 52 (47.27%) | 15 (13.64%) | 0 (0%) | 15 (13.64%) | 45 (40.91%) |
| <i>Voor</i> | 73 | 31 (42.47%) | 15 (20.55%) | 0 (0%) | 15 (20.55%) | 28 (38.36%) |
| Total | 619 | 225 (36.35%) | 111 (17.93%) | 13 (2.10%) | 98 (15.83%) | 289 (46.69%) |

2.3 Predicate frequency

Infrequent predicates are not suited for the analysis in the first study of this thesis. A quantitative approach such as the one used, does not work well if there is only a small set of sentences available for a certain predicate. For this reason, predicate frequencies were needed in order to decide which predicates to use, and which not to. As I intended to use the same set of predicates throughout all studies, a predicate that can be used for one study but is unsuitable for another will not be included at all. From the 619 predicates, 289 were unproblematic in the annotated categories altogether. For each of those, predicate frequencies were looked up.

The Dutch word frequency databases one would normally turn to (e.g., SUBTLEX-NL) do not allow for searching for preposition-verb combinations—in particular those combinations that are not always adjacent to each other. Instead we used the web application PaQu (Parse and Query; Odijk, van Noord, Kleiweg, & Tjong Kim Sang, 2017), which allows a user to search for “dependency triplets” in (provided or uploaded) text corpora with annotations on the syntactic structure. Fortunately, the web application PaQu (Parse and Query) was developed with a large focus on this purpose (Odijk, van Noord, Kleiweg, & Tjong Kim Sang, 2017). Using their online accessible interface, you can search a number of corpora already included, or any corpus you wish to upload yourself, for “dependency triplets”. Such triplets consist of a combination of a head word, a dependent word, and a syntactic relationship between the two (parts-of-speech can be specified for both of the words). We used the relationship “prepositional complement” or “pc” to extract frequency information of the verb-preposition combinations that were left from the first selection round, specifying the respective parts of speech (‘verb’ and ‘preposition’). I utilised a selection criterium of at least 100 hits needed for each predicate, measured in number of sentences found, so there would be enough data available to perform the analyses in the first study in this thesis. Analyses were performed on the *WR-P-P-G* part of the *Lassy Groot* corpus, which consists of newspaper articles (Van Noord, Schuurman, & Vandeghinste, 2006; Van Noord, 2009). It contains a total of 14,974,458 sentences, which were syntactically annotated by the automatic parser *Alpino* (Bouma, van Noord, & Maloud, 2001).

The frequency numbers showed that 187 predicates had at least 100 hits in the corpus, while 34 predicates did not occur in it altogether. The average number of hits was 1869.79, ranging from 0 to 44,948. Table 2.3 shows the frequency numbers split out for each separate group.

Table 2.3

An overview of the predicate frequencies, both per predicate group and for all predicates in total.

| Preposition | Average number of hits | Minimum | Maximum | Number of predicates with more than 100 hits |
|--------------|------------------------|---------|---------|--|
| <i>Aan</i> | 3193 | 0 | 22260 | 30 |
| <i>In</i> | 1670 | 0 | 12946 | 10 |
| <i>Naar</i> | 1271 | 0 | 10231 | 14 |
| <i>Om</i> | 702 | 0 | 7352 | 9 |
| <i>Op</i> | 1532 | 0 | 16089 | 25 |
| <i>Over</i> | 1175 | 0 | 12685 | 18 |
| <i>Tegen</i> | 625 | 0 | 3107 | 11 |
| <i>Tot</i> | 1911 | 0 | 20209 | 15 |
| <i>Van</i> | 2111 | 0 | 13469 | 33 |
| <i>Voor</i> | 3295 | 0 | 44948 | 22 |
| Total | 1869 | 0 | 44948 | 187 |

The ten most frequent predicates from each group were selected for the final set that was used in all studies in this thesis. As Table 2.3 shows, however, not all groups contain ten predicates with at least one hundred hits in the corpus. The group that governs *om* only contains nine such predicates, as the tenth most frequent predicate only had thirty hits in the corpus. This left me with a total of 99 predicates for this thesis. A complete list of these predicates, as well as the number of hits for each predicate is included in Appendix A.

2.4 Loan words

All predicates were checked for a potential status as loan word. This is important as loan words, especially when borrowed between the languages of interest, would interfere with the cross-linguistic comparison in the second study. In order to prevent problems, etymologies were checked for all predicates/translations. How translations were chosen for each of the predicates is detailed in section 5.2, wherein the methodology of the cross-linguistic study is detailed.

For the Dutch predicates, the website *Etymologiebank* was consulted (Van der Sijs, 2010). The website combines a large number of already available (etymology) dictionaries in one. Information is incorporated from over sixty dictionaries and covers 100,000 unique entries. A few of the predicates either do not have an entry on the website, or do have an entry, but only information on the meaning is provided. Most of these concern complex predicates, made up of one or more prefixes attached to the root. If this happened, I instead looked up the etymology of the root. For most of the predicates, the etymological data from *Etymologiebank* showed a Middle Dutch (13th to 15th century), or even Old Dutch (6th to 12th century) etymon. For the rest, it showed a Early Modern Dutch (16th and 17th century) etymon.

The etymology of all English predicates was checked in the online version of the *Oxford English Dictionary* (Oxford University Press, n.d.), which provides extensive information on word origins. Most translations in the data were either inherited from Germanic, or came

(sometimes via French) from a Latin origin. None though, were recently borrowed, and certainly not from either of the other languages in the cross-linguistic study.

For the German predicates, the online available dictionary *Duden* was used ('Duden online', n.d.). For most words, this dictionary provided etymological information, giving the Middle (High) German (11th to 15th century) and sometimes Old High German (mid 8th to 11th century) etymon. However, for twenty of the predicates in the selection, no etymological information was listed. This was often for the same reason as with the Dutch predicates: they were complex verbs. One such example from the set is *einbeziehen* 'include', which is made up of the verb *ziehen* 'pull' and the two prefixes *ein-* and *be-*. Unlike with Dutch, however, my limited knowledge of the German language disallowed me to divide these verbs and extract the root myself. To that end, I looked up the remaining predicates in the *German Wiktionary* ('Wikiwörterbuch: ein Wiki-basiertes freies Wörterbuch', n.d.). This reference work, unlike the *Duden*, provides data on how these verbs were composed, giving the information needed to look up the etymology of the root. In the end, none of the German translations were (recently) borrowed.

To conclude, none of the predicates selected based on their frequencies needed to be removed from the dataset and replaced with another. The final set of 99 predicates, along with frequency measures for each predicate can be found in Appendix A. Etymological information is not included in this list, as data from all three sources (*Etymologiebank*, *Oxford English Dictionary* and *Duden*) is copyrighted and is therefore prohibited from being copied, published or otherwise shared for non-personal uses. However, both the *Etymologiebank* and *Duden* are available for use freely via the websites listed in the references section of this thesis. The *Oxford English Dictionary* is available with a subscription, such as the one offered through the Radboud University Library.

Chapter 3. Semantic similarity of Dutch corpus data

3.1. Introduction

Firth (1957, p. 11) popularised the idea that “You shall know a word by the company it keeps!” He explains that part of the meaning of a word, e.g. *cows* can be found in its use in sentences in which it is most characteristically embedded, such as *they are milking the cows*. This idea can be further extended to the notion that words that are semantically similar to each other will occur in similar texts (i.e., Harris’ *distributional hypothesis* (1954)). For example, you can imagine that texts about a *dog* or a *puppy* will use much of the same words, while texts that talk about a *spaceship* will use very different words. *Dog* and *puppy* might also very well be used in the same sentence, but semantic similarity goes beyond two words’ co-occurrence. Kwantes (2005) gives the example of *general* and *admiral*. These words clearly have some semantic similarity: they are both ranks in the military. However, admirals are specific to the navy, while generals exist in the other military branches. These terms will therefore only rarely co-occur in a text. They will, however, both be used in the vicinity of words such as *military* and *officer*. Looking at the context in which two words occur will therefore provide a great amount of information about their semantic similarity. The aim of this chapter is to use this theory to build a semantic model of the predicates under investigation, and compare them to each other.

An often-used way of carrying out such an analysis, is a method called Latent Semantic Analysis (LSA; Deerwester, Dumais, Furnam Landauer, & Harshman, 1990; Landauer & Dumais, 1997; Landauer, 2007; Kintsch, McNamara, Dennis, & Landauer, 2007). It is a mathematical technique that is used in Natural Language Processing, specifically distributional semantics.³ It analyses the relationships between sets of documents and the terms they contain. As Koeman & Rea (2014, p. 7) describe, “LSA extracts what appears to humans to be semantic meaning.” It can be applied, for example: 1) to find documents on similar topic (even across languages); 2) to provide better results in search engines, as it allows the system to not only produce results that match the exact words in the query, but also synonyms/semantically related terms; 3) to build a topic model of a given document, or even extract text summaries.

Essentially, LSA takes a text, makes a list of all words that occur in that text and their frequency, and compares that with its list of another text. It can then group texts that contain similar words, and, therefore, have a similar semantic context.

In this study, Dutch corpus data will be used for each of 99 predicates under investigate, and LSA will be used to calculate similarity scores for each pair of predicates. If two predicates are semantically similar, the texts they occur in will be found to be similar in an LSA analysis. Since I am investigating the hypothesis that the choice of preposition is not arbitrary, but instead related to the preposition of a predicate, I expect to find that semantically similar groups of predicates share the same preposition.

³ It is also used in the field of information retrieval, where it is often referred to as “Latent Semantic Indexing” (LSI). Otherwise, these terms are often used interchangeably in literature.

3.2. Data

For this study, the web application PaQu was used to extract corpus data (as described in section 2.3) from the *WR-P-P-G* part of the *Lassy Groot* corpus (van Noord, Schuurman, & Vandeghinste, 2006), one of the biggest corpora available in the search application. A total of 309,018 sentences were collected containing data on the 99 predicates under investigation.

The data was structured as follows: 99 data files were downloaded from PaQu, one for each predicate. Each hit—a sentence containing the verb and preposition in it—started on a new line. Each line started with meta data about which section of the corpus it came from, which was removed before analysis. The rest of the lines consisted of the sentence itself, in plain text.

Secondly, the preposition of the predicate in question was removed from each sentence. This to remove an unwanted bias wherein predicates from the same preposition group would inherently be more similar to each other than to predicates from a different preposition group, as the overlapping context would be larger. All sentences with more than one occurrence of the preposition in question were set apart. These sentences could either be like sentence (22), in which both prepositions were governed by the verb, and thus instances of the predicate under investigation, or like sentence (23), in which only the first *aan* is governed by *danken*, while the second *aan* is governed by *deelgenomen* ‘participated’. In the former case, both instances of *aan* should be removed, as they are both part of the predicate under investigation; in the latter case, only the first instance of *aan* should be removed.

- (22) *De huidige ontwikkeling van de EU is **aan** iets anders te **danken** dan **aan** het overwinnen van tegenstellingen.*
“The current development of the EU is **due to** something other than [**to**] overcoming contradictions.”
- (23) *De groei is te **danken** **aan** de deelname van veteranen die **aan** vredesmissies hebben deelgenomen.*
“The growth is **due to** the participation of veterans who have participated **in** peacekeeping missions.”

It could not be established automatically whether both of the prepositions were governed by the verb, or only one of them was. As the number of sentences with a double preposition (i.e., 38,372) was too large to annotate manually, only those sentences were included in this study that contained a single occurrence of the preposition in question. After removing these sentences, 270,646 sentences were left, and all predicate files still contained more than the lower limit of 100 sentences set during predicate selection. An overview of the total number of sentences for each predicate, as well as the number of sentences removed and kept is attached in Appendix B.

With this cleaned-up data, further preparations could be made for LSA. Punctuation was stripped from the text, as well as capitals and diacritics. Other than these, Gefen et al.’s (2017) *guide to text analysis with LSA* mentions three transformations that are often used to prepare the data for LSA: stemming, lemmatisation, and stop word removal.

Stemming is the process of removing inflection from words, and reducing them to their word *stem*. This is not necessarily the same as the root of a word, nor is it always a form of the word that exists in actual language use. For example, the Natural Language toolkit

(NLTK; Bird, Klein, & Loper, 2009) PorterStemmer (Porter, 1980) reduces *language* and *languages* both to “*languag*”, and *argue*, *arguing* and *argued* are reduced to “*argu*”. This process decreases the size of the dictionary (all words used in the corpus) and is often used in the field of machine learning.

Lemmatization is similar to stemming, in that it is also a normalisation technique. It is the process of identifying the lemma of a word and, using the part of speech, bringing that word down to its base form. For a great number of words, stemming and lemmatization give the same result. The major difference between the two processes, however, is that a lemma is always a meaningful base form, while the word stem can, as described above, be meaningless on its own and be incorrectly spelled. This means, that in some cases, stemming and lemmatization provide different results. Stemming the words *caring* and *stripes* would give the word stems *car* and *strip*, while lemmatization would give the lemmas *care* and *stripe*.

Finally, stop words are words that do not bring much meaning to a sentence. These are some of the most common words in a language, and are often function words. English examples are *the*, *I*, and *is*. Removing these words would bring a larger focus to the content words of a text.

It appears that opinions vary greatly about whether or not to use such transformations. The creators of LSA themselves advise against it (Landauer & Dumais, 2008). They state that “[i]n keeping with the underlying theory and model, neither stemming nor stop-listing is appropriate or usually effective”. These stop words actually attribute to the meaning of natural language, and meaning cannot be accurately be understood without them. Schofield and Mimno (2016) also speak to the effectiveness of stemming, in a practical study. Their investigation shows that stemming does not improve coherence, but might actually reduce predictive likelihood. They point out that stemming can very well be useful when dealing with a corpus containing heavily varied word forms—such as a corpus they used consisting of reviews from the crowd-sourced review website *Yelp*—but as the present corpus consists of newspaper articles, that is not the case. Landauer and Dumais (2008) do acknowledge that stemming and lemmatization can be useful when LSA is used to compare word strings shorter than average paragraphs. That, however, is not the case here, as I am comparing whole documents of text (the fact that each document consists of individual sentences should not factor into this). According to these papers, the transformations mentioned by Gefen et al. (2017) are not required, and could actually negatively affect the results of the analysis. Therefore, I decided not to apply these transformations in the initial analysis.

3.3. Analysis

The Python library *Gensim* was used to perform the analysis (Řehůřek & Sojka, 2010). A dictionary was built, and the class *LSIModel* from the module *models* was applied to create the LSA space. Each pair of predicates was then compared using the class *Similarity* from the module *similarity*, which calculates the *cosine similarity*. This is a similarity measure between 0 and 1. The closer the cosine is to 1, the more similar two documents are. One cosine similarity score does not mean anything in isolation, but it can be used to compare if item A is more similar to item B or to item C.

One of the difficulties in LSA is in determining the number of dimensions to use. *Gensim* defaults the number of dimensions to 200, but literature very much disagrees on the optimum number to use. For example, in a paper on the effects of varying dimensionality, Bradford (2008) mentions that for a five million document corpus around 400 dimensions

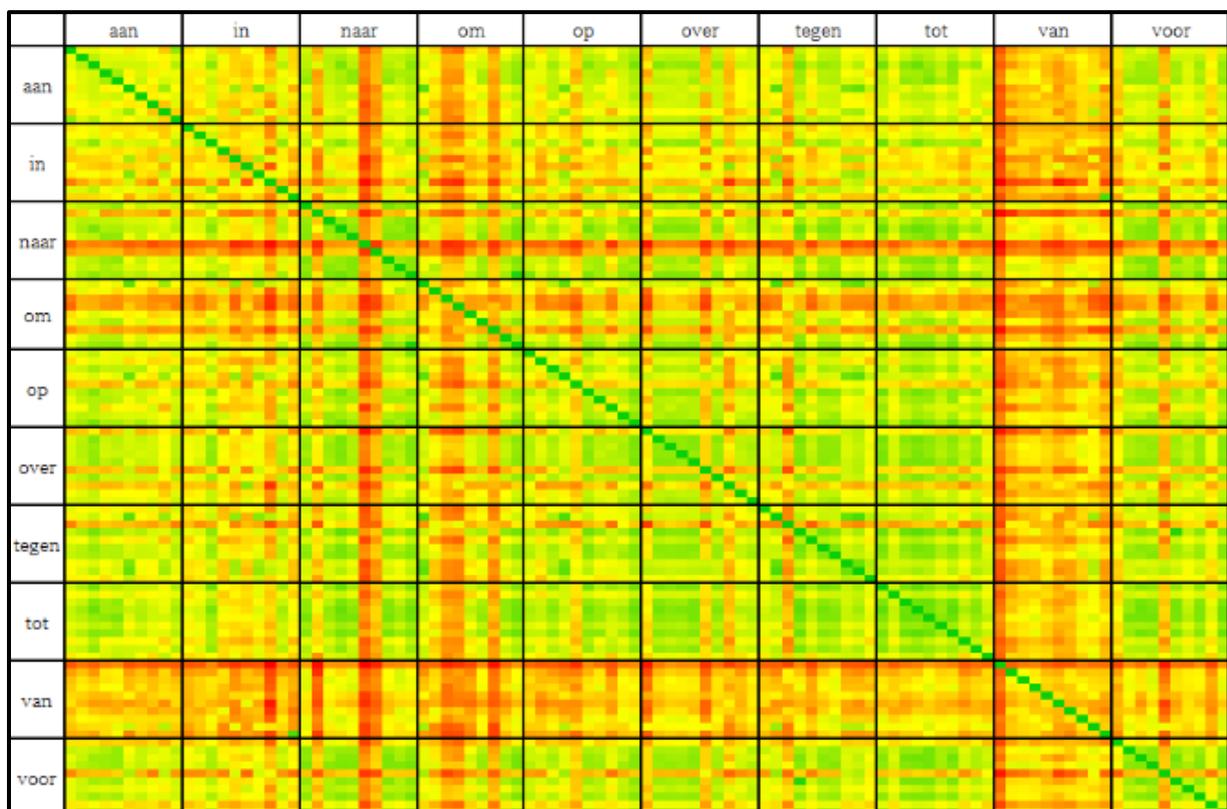
provides the best performing model, and that there is an ‘island of stability’ between 300 and 500 dimensions. Landauer and Dumais (2008) agree that around 300 dimensions is usually best, but they also state that somewhere between 50 and 1000 dimensions is optimal for most studies. Landauer (2002) places that interval between 100 and 1500. One researcher even found that for some tasks, accuracy increases with the number of dimensions (Laham, p.c. in Quesada, 2007). Quesada (2007) states that the optimal dimensionality is size dependent, because big and small spaces have very different characteristics.

Because there is such disagreement on an optimal number, or a range in which the optimum falls, I ran the LSA several times, with a different number of dimensions specified each time: 100, 200, 300, 1000, and 1500 dimensions. The models resulted in matrices with no visual differences, so only the one with 300 dimensions is included in this thesis. As for the other parameters: the corpus was input as nested bag-of-words list for each file, and the variable *id2word* was set to ‘dictionary’.

Figure 3.1 presents the whole matrix of cosine similarities between each pair of predicates. A colour scale is applied to the matrix to allow for a visual inspection of the data. As stated before, cosine similarity scores have no interpretation out of context, so the colours are not attached to set scores. Rather, the lowest occurring score is red, while the highest occurring score is green.

Figure 3.1

The full LSA matrix (300 dimensions) with a colour scale applied. Green = most similar, red = least similar.



The highest occurring score is reserved for the cells on the top-left to bottom-right diagonal. This only means that the model works correctly, because a predicate compared to itself should always be most similar. Other than that, though, there is no clear pattern to the data. The only other observation that can be made, is that it appears as though the predicates that govern *van* are a little less similar to other predicates than the rest. However, they are also similarly dissimilar to other predicates from the same group.

In order to see which words have the greatest impact on the model, and thus the cosine similarity scores, a key word analysis was carried out. Using the class *keywords* from Gensim's *summarization* module, the top five key words from all files were listed.⁴ These are the five key words with the highest score, which means the program has extracted these as the most important to or representative of the meaning of the text as a whole. Table 3.1 lists, to portray what kind of words are listed, the key words for the first ten predicates (all predicates with *aan*). A table with the full set of key words can be found in appendix C.

Table 3.1

The top five key words from the first ten files.

| Predicate | <i>kw1</i> | <i>kw2</i> | <i>kw3</i> | <i>kw4</i> | <i>kw5</i> |
|------------------------|------------|------------|-------------|------------|------------|
| <i>danken_aan</i> | van | het | dat | die | eens |
| <i>hechten_aan</i> | van het | eens | hecht | dat | |
| <i>liggen_aan</i> | van | het | ligt | eens | die |
| <i>lijden_aan</i> | eens | van | lijdt | het | die |
| <i>onderwerpen_aan</i> | eens | van het | onderworpen | die | |
| <i>storen_aan</i> | van | het | een | die | zich |
| <i>verbinden_aan</i> | van | eens | het | verbonden | zijne |
| <i>voldoen_aan</i> | van | voldoen | het | eens | die |
| <i>wagen_aan</i> | eens | van | zich | het | waagt |
| <i>wijten_aan</i> | van | het | dat | eens | die |

First, this set of key words provides insight in the dissimilar nature of the group of *van* predicates observed earlier. The word *van* is present in a significant number of these lists of top key words (it is listed 82 times in isolation, and makes up 18.6% of all key words). Despite that, it is represented among exactly none of the top key words for the group of predicates that govern *van*. That can be easily explained. In each sentence in these files, the word *van* was removed, as it belongs to the predicate. If a sentence contained a second *van*, there was no way of automatically determining if it belonged with the predicate or not, so all those sentences were discarded. The number of sentences for which that was the case was also a lot higher for *van* predicates than for other predicates (35.24% removed for *van* predicates vs. 12.02% in total). If, for instance, it were not *van* that was so prominent in all files but *in*, we would see a similarity matrix in which predicates that govern *in* were more dissimilar to other verbs.

Secondly, you can see that this table includes a lot of function words. In fact, function words make up 60% of all top five key words (287 out of the total 478 key words provided). Additionally, the word *eens* occurs 91 times in the list (19%). It seems, therefore, that these

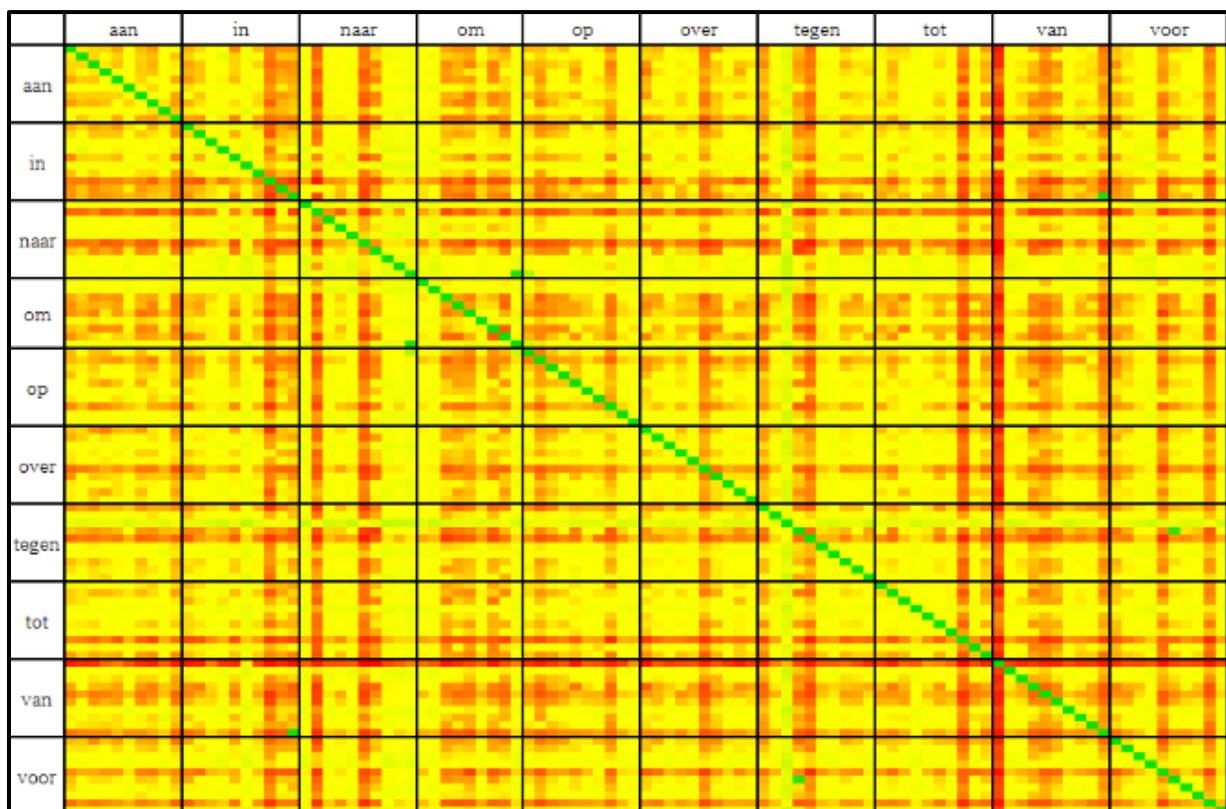
⁴ In some cases, a key word actually consisted of a combination of two words (most often *van het* 'of/from it'). In these cases, only the top four key words were provided by the program.

words have to big an influence on the text. For that reason, I ran the analysis again, this time following the original advice from Gefen et al. (2017) to use stop word removal to the data. All stop words were removed using the list of Dutch stop words from NLTK (Bird, Klein, & Loper, 2009). Afterwards, the LSA was run again, using the same parameters as described before.

Again, there was a high degree of similarity between the matrices with the different numbers of dimensions, so the matrix that resulted from running the analysis with 300 dimension is displayed in Figure 3.2.

Figure 3.2

The full LSA matrix (300 dimensions) after stop word removal with a colour scale applied; green = most similar, red = least similar.



This LSA matrix looks very similar to the previous one, in that the only clear pattern to the data is in the main diagonal. There are small number of bright green cells outside this diagonal, namely the scores for the predicate pairs *pleiten voor – pleiten tegen*, *voorzien van – voorzien in*, and *vragen naar – vragen om*. This can be explained by the fact that they share the verb.

A second key word analysis was carried out on the data after the stop words were removed. Table 3.2 lists the top five key words from the first ten predicates, as well as the key words for the six predicates mentioned above. The table with the complete set of key words can be found in appendix C.

Table 3.2

The top five key words from the first ten files after stop word removal, along with the top five key words for *pleiten tegen*, *pleiten voor*, *voorzien in*, *voorzien van*, *vragen naar*, and *vragen om*.

| <i>Predicate</i> | <i>kw1</i> | <i>kw2</i> | <i>kw3</i> | <i>kw4</i> | <i>kw5</i> |
|------------------------|-------------|-------------|------------|------------|--------------|
| <i>danken_aan</i> | danken | dankte | vooral | volgen | succes |
| <i>hechten_aan</i> | hecht | hechten | belang | gehecht | waard |
| <i>liggen_aan</i> | ligt | lag | basis | liggen | volgen |
| <i>lijden_aan</i> | lijdt | lijden | leed | mensen | ziektes |
| <i>onderwerpen_aan</i> | onderworpen | onderwerpen | werden | onderwierp | onderzoekers |
| <i>storen_aan</i> | stoort | storen | stoorde | stoor | gestoord |
| <i>verbinden_aan</i> | verbonden | verbinden | verbindt | naam | wel |
| <i>voldoen_aan</i> | voldoen | moeten | eisen | alle | nieuws |
| <i>wagen_aan</i> | waagt | wagen | gewaagd | waagde | eerst |
| <i>wijten_aan</i> | wijten | wijt | volgen | vooral | deels |
| <i>pleiten_tegen</i> | pleit | pleiten | pleitte | gepleit | vooral |
| <i>pleiten_voor</i> | pleit | pleiten | pleitte | gepleit | nieuws |
| <i>voorzien_in</i> | voorzien | nieuw | eigen | moeten | behoefte |
| <i>voorzien_van</i> | voorzien | nieuws | alle | twee | warden |
| <i>vragen_naar</i> | vragen | vroege | gevraagde | vraagt | vraag |
| <i>vragen_om</i> | vraagt | vragen | gevraagde | vroege | vroegen |

Removing the stop words did away with the many function words that dominated the previous set and that there are considerably more content words in this set. However, a large number of these content words are various forms of the verbs themselves. There is actually only a single predicate where the verb is not listed in the key words, and that is the predicate *kunnen tegen*. That makes sense, because *kan*, *kunnen* and *kon* are all in the list of stop words that were removed from the data, and these are all forms of the verb *kunnen*;⁵ the first two are also, by far, the most frequent forms, according to data from SUBTLEX-NL (Keuleers, Brysbaert, & New, 2010). All other predicates list their verb at least once, and fifteen of them actually only have different forms of the verb listed as their key words. Before stop word removal, 98 (20.5%) of the key words were the verb itself. That number has increased to 310 (63.7%). So while we've made progress in getting rid of function words, we've now created a new problem. Because such a large part of the key words is made up of the verbs themselves, there is still little comparison being made between the actual content of the different files. In order to make a meaningful comparison between the files, the verb itself needs to be removed from the data. However, this would require a list of all the differently inflected versions of all verbs, which was unavailable. Therefore, this step is not feasible in the scope of this thesis and will not be taken. The discussion will list recommendations for further analysis.

⁵ Only missing are *kunt* (a more formal alternative of *kan* in simple present, second person singular) and *konden* (simple past, plural).

3.4. Discussion

This study set out to build a semantic model of the predicates in this thesis. Using Latent Semantic Analysis, a model was created to compare overlapping contexts as a way of calculating similarity between pairs of predicates. I expected the data to show that predicates within a preposition group are more similar to each other than to predicates that govern a different preposition. However, neither the first nor the second model described reflected such a pattern.

The first model was constructed with only minor transformations to the data. Punctuation, capitalisation and diacritics were removed; stop word removal and stemming or lemmatisation were not used, as several researchers advised against it. The similarity matrix that resulted from this model showed no clear patterns. A key word analysis showed that function words had a very large effect on the model. For this reason, stop words were removed, and a second model was constructed. Although this model was based largely on content words, key word analysis revealed a large number of these words to be the verb itself. This resulted in all files being quite different from each other, as they shared almost none of the most prominent key words. Again, no pattern was found in the data.

However, I do not consider this lack of result to mean that there is no such trend in the corpus data. Rather, I believe that this study was limited by a data preparation that was not extensive enough. Stop word removal moved us closer towards a semantic model based on content words instead of function words, but it did not eliminate the problem of the verb itself being one of the main influences on the model. The analysis would focus more on the semantics of the context of verbs, when the verbs themselves are removed from the data. Additionally, it would likely give more weight to content words, and nouns in particular, which could lead to a better understanding of the semantic domain of a predicate. Because stop word removal turned out to be of positive influence to the data set, I also believe other recommendations from Gefen et al. (2017) should be followed. However, this is not feasible within the time frame of this thesis, and shall have to be left for new studies.

Further research should be done to investigate the hypothesis this study presented. The steps I have identified that still need to be taken are stemming or lemmatisation, and removing all different variations of each of the verbs (in their own files). This would allow for a model to be created that is more focused on the context of the predicates than the predicates themselves. It is very much still possible that this model would show increased similarity among predicates that govern the same preposition.

Chapter 4. Cross-linguistic comparison

4.1. Introduction

This chapter will describe the results of a cross-linguistic investigation of prepositional predicates. For this, I wanted to compare Dutch with other Germanic languages. I selected German and English, as Dutch holds a sort of ‘middle ground’ between the two languages (van Haeringen, 1956).

The aim of this study was to compare overlapping prepositions across languages, with the underlying idea that if two verbs use the same preposition in language 1, and also share their preposition in language 2, they are more similar than two verbs that use the same preposition in language 1, but different ones in language 2. To illustrate, in the example below, pair A (V_1 and V_2) uses the same preposition in both Dutch and English, while pair B (V_3 and V_4) does so in Dutch, but not in English. Therefore, pair A would be more similar than pair B. It is important to note that these prepositions need not be direct translations of each other, the relevant thing is corresponding prepositions within a language.

- A. NL: V_1 -aan, V_2 -aan
EN: V_1 -to, V_2 -to
- B. NL: V_3 -aan, V_4 -aan
EN: V_3 -to, V_4 -on

If you would compare each pair of predicates in a set in such a manner, you could create a semantic map (as described in Wälchli, 2010, and Wälchli & Cysouw, 2012). Such a semantic map is a method of visually representing cross-linguistic data in a (most often) two-dimensional graph. This type of figure allows for an easy comparison and categorisation of the predicates. As Wälchli and Cysouw (2012) describe, semantic maps have the benefit of being able to both look at the big picture and see general tendencies, but still retain the small details in order to identify specific features. First created as a method to study patterns of polysemy and co-expression, it is now used for all types of data. The main idea is that the spatial arrangement of the data reflects (dis)similarity in the data (Anderson, 1982).

In this chapter, I therefore aim to create a semantic map of the predicates in this study, using translation data from German and English for a dissimilarity score. As I suspect that the process of selecting a preposition to complement a verb is not random, I expect this study to have the following results: where Dutch predicates use the same preposition, their English and German translations will do the same, and therefore group together in the semantic map.

This chapter will first discuss how the translations were realised, the creation of a dissimilarity matrix, and finally analyse the data in order provide insight into my hypothesis. An in-depth look into the translations will investigate whether the semantics of the preposition are at play during selection.

4.2. Data

4.2.1 Translations

The first step of this study was to find equivalent translations for all verbs. Section 2.4 already describes some etymological data on the translations. This section will explain in more detail how these translations were selected.

First, an online version of the Dutch dictionary *Van Dale* was used to collect all possible English and German translations for each predicate, along with any prepositions, and, for German, any cases mentioned alongside them — though neither was provided often, in either language. Based on the example sentences, any extra information provided with the individual words (e.g., “only for human referents”), and my own intuition, I chose a main English translation and a main German translation for each predicate. Often, most—or even all—possible translations used the same preposition. Take, for instance, *zich baseren op*, which has four translations listed by *Van Dale*: *base (oneself) on*, *found on*, *rely on*, and *go on*. In this case, choosing one translation or another would have had no influence on the following analysis, as that is solely based on the preposition governed (though care was still taken in selecting a particular translation). However, with a few predicates, two translated prepositions vied for the spotlight. This was, for example, the case with *zich bekommeren om*, whose possible translations were divided amongst governing *about* and *with*: *worry about*, *bother about*, *trouble oneself about*, *concern oneself with* and *bother oneself with*. When no clear preference was found for either preposition, both translations were chosen, in this case one with *about*, and one with *with*. Additionally, in most cases, choosing the translation best representing the Dutch meaning was no problem, but for a few predicates, no single translation fully covered the meaning. In those instances, the two translations were chosen that covered the meaning most extensively together. In total, six Dutch predicates were paired with two translations in English, and one predicate with two translations in German.

After that, I used the *Oxford English Dictionary* (Oxford University Press, n.d.) and the *Duden* (‘Duden online’, n.d.) to check, respectively, all English translations and all German translations. I checked to see whether the translations were used in similar sentences as the Dutch equivalent would be. If a preposition had been mentioned in the *Van Dale*, I confirmed it in the dictionary; if none was provided, I used the dictionary to find a preposition, or lack thereof. Also, as described in section 2.4, these reference works were used to find out etymological information on all prepositions.

For German, there was the extra step of checking the case marking. For a few predicates, this information was provided within the *Van Dale*, in which case it was checked in the *Duden*. If not, the case marking of the object was looked up therein. If neither dictionary explicitly mentioned the case marking, there were two possible routes, depending on the type of preposition. German prepositions can be divided into two groups: one-way prepositions and two-way prepositions (or, in German, *Wechselpräpositionen*). One-way prepositions are prepositions that always govern the same case, may that be genitive, dative, or accusative. Two-way prepositions, on the other hand, can change case between dative and accusative, based on the intended meaning. As Zwarts (2005) finds, there are clear divisions to be made between the type of prepositions that govern the dative and prepositions that govern the accusative case. The former group consists mainly of locative or source prepositions, while the latter consists of route or goal prepositions. This is also transferrable to two-way prepositions, where a locative meaning would see it used with a dative case marking, and directed motion with an accusative case marking.

A list of prepositions can be found in Table 4.1, sorted based on the case they govern (‘German/Grammar/Prepositions and Postpositions’, 2020). Only those prepositions are listed that occur will occur in the dataset.

Table 4.1

German prepositions⁶ and the case they govern

| Genitive | Dative | Accusative | Dative/Accusative (two-way) |
|----------|--------|------------|--------------------------------|
| / | aus | für | an |
| | bei | gegen | auf |
| | mit | um | in |
| | nach | | über |
| | von | | vor |
| | zu | | |

If a one-way prepositions was used with a certain translation, one need only check this list to see which case should accompany it. If a two-way preposition was used, the example sentences in *Duden* were checked to see the case marking on the object. If no preposition was used, and no case marking information was provided, this latter method was used as well.

All English translations were looked over by both my thesis supervisors, and the German translations by a native German speaker, who is also highly fluent in Dutch. This resulted in some alternative translations being chosen instead of the original selection, or an additional translation being added to better embody the multiple facets of a predicate’s meaning. The full set of translations for both languages is enclosed in Appendix D.

4.2.2 Dissimilarities

In order to make a systematic comparison, the data was transformed into a dissimilarity matrix. To make this dissimilarity matrix, Hamming distances were calculated between each pair of predicates. A Hamming distance is a metric that measures the number of positions where two items of equal length are different. These may be, among other things, letters in a word, digits in a number, or items in a list. For instance, the word pair *{bed, bet}* is different in one position (marked in bold): the position of the third letter. Therefore, the distance between the two words is 1. The number pair *{111, 123}* is different in two positions, and thus the distance is 2. A similar analysis may be applied to the lists of translations we have.

First, only the prepositions and, for German, case marking are of interest, and are isolated. For Dutch and English, this was rather straightforward. Either a preposition was copied, or a \emptyset marking was copied if the translation had a non-prepositional complement. For German, there were a few more things to consider. If the translation had a one-way preposition, only the preposition was copied, as any case information is redundant with these types of prepositions. If the translation had a two-way preposition, the case it governed was copied along with it. If the translation had no preposition, \emptyset marking was copied, along with

⁶ The term *prepositions* is here used in its broader sense and includes postpositions as well, as some German adpositions can occur after their complement (e.g. *dem Feind entgegen* ‘against the enemy’ (example borrowed from Zwarts, 2005, p. 3)).

the case of the object, thus treating it the same way as a preposition that can govern more than one case. [more info here on whether or not the case matters with zero-preposition marking]

This then leaves us with the information displayed in the columns *predicates* to *DE* of Table 4.2 (which contains the first ten predicates of the data set). The final column of the table contains the lists that are composed of the before mentioned information with which the Hamming distances are calculated.

As in the examples before, each item—in this case, instead of letters in a word or digits in a number, items in a list—in list A is compared to the item in the corresponding position in list B. As you can see in Table 4.2, predicates 1 and 7, $\{[aan, to, dat], [aan, to, mit]\}$, have a Hamming distance of 1, as only the German preposition differs between the two.

Table 2.2

A subset of the first ten Dutch predicates with the corresponding translated prepositions and cases, and the lists the Hamming distances are calculated with.

| # | Predicates | NL | EN | DE | list |
|----|-----------------|-----|------|------------|-----------------------|
| 1 | danken_aan | aan | to | dat | [aan, to, dat] |
| 2 | hechten_aan | aan | to | an-dat | [aan, to, an-dat] |
| 3 | liggen_aan | aan | on | an-dat/bei | [aan, on, an-dat/bei] |
| 4 | lijden_aan | aan | from | an-dat | [aan, from, an-dat] |
| 5 | onderwerpen_aan | aan | to | dat | [aan, to, dat] |
| 6 | storen_aan | aan | Ø | an-dat | [aan, Ø, an-dat] |
| 7 | verbinden_aan | aan | to | mit | [aan, to, mit] |
| 8 | voldoen_aan | aan | Ø | acc | [aan, Ø, acc] |
| 9 | wagen_aan | aan | Ø | an-acc | [aan, Ø, an-acc] |
| 10 | wijten_aan | aan | on | dat | [aan, on, dat] |

If a predicate in list A had multiple translations in one or both of the languages, each translation was compared against the corresponding translation in list B, and each comparison counted for half a point for the Hamming distance. So, if neither preposition in list A matched the preposition in list B, the Hamming distance is 1. If one matched, the Hamming distance would be 0.5, and if both matched, the distance would be 0. For example, predicates 2 and 3 differ in their English translations, and in one of the German translations of list B, $\{[aan, to, an-dat], [aan, on, an-dat/bei]\}$, and thus have a Hamming distance of 1.5.

In the occasion that the German translation consisted of a preposition-case combination (which can be the case only for two-way prepositions, and for non-prepositional complements), first the prepositions were compared. If they didn't match, the Hamming distance would be 1. If they matched, case marking was subsequently compared. If both matched, the Hamming distance would be 0. If only the preposition matched, but not the case marking, the Hamming distance would be 0.5. Table 4.3 shows the different types of combinations and the corresponding Hamming distance. The last three rows of the table contain preposition-case combinations, which are specifically for certain cases with German translations.

Table 4.3

Hamming distances for different combinations, where P = preposition and C = case.

| List A | List B | Hamming distance |
|--------------------------------|-------------------------------|------------------|
| P ₁ | P ₁ | 0 |
| P ₁ | P ₂ | 1 |
| P ₁ /P ₂ | P ₁ | 0.5 |
| P ₁ C ₁ | P ₁ C ₁ | 0 |
| P ₁ C ₁ | P ₁ C ₂ | 0.5 |
| P ₁ C ₁ | P ₂ C ₁ | 1 |

Keeping with the examples from Table 4.2: Predicates 6 and 9 have the same German preposition, but a different case marking, so this combination has a Hamming distance of 0.5, $\{[aan, 0, an-dat], [aan, 0, an-acc]\}$.

The dissimilarity matrix was calculated automatically by a Python program. To portray, a small section of the dissimilarity matrix is displayed in Table 4.4. As you can see, the main diagonal (from top-left to bottom-right) is always 0, as a predicate compared to itself cannot have any dissimilarities. Additionally, because this matrix has the same items in the rows and columns, the parts below and above the diagonal are mirror images of each other.

Table 4.4

Sample of the dissimilarity matrix of the first four predicates.

| | <i>danken_aan</i> | <i>hechten_aan</i> | <i>liggen_aan</i> | <i>lijden_aan</i> |
|--------------------|-------------------|--------------------|-------------------|-------------------|
| <i>danken_aan</i> | 0 | 1 | 2 | 2 |
| <i>hechten_aan</i> | 1 | 0 | 1.5 | 1 |
| <i>liggen_aan</i> | 2 | 1.5 | 0 | 1.5 |
| <i>lijden_aan</i> | 2 | 1 | 1.5 | 0 |

4.3. Analysis

4.3.1 Multidimensional Scaling

Following the example of Wälchli (2010) and Wälchli and Cysouw (2012), I wanted to explore the dissimilarity data using Multidimensional Scaling (MDS). MDS is a data visualisation method that takes multidimensional (in this case 99-dimensional) dissimilarity data among pairs of objects, and creates a low-dimensional representation of it (Borg & Groenen, 2005). Most researchers use it to create a two-dimensional scatterplot of dissimilarity data. Pairs with a low dissimilarity score would be close together in such a scatterplot, while pairs with a high dissimilarity score would be far apart. This allows you to explore the structures in your data set visually, and observe grouping among the different datapoints.

With MDS comes the goodness-of-fit measure *stress*, the residual sum of squares, which measures how well a given configuration fits the data (Kruskal & Wish, 1978; Kruskal, 1964).

It can be described as the “error of representation” (Borg & Groenen, 2005, p. 41), and for this reason, Borg and Groenen adopt the more fitting term “badness-of-fit” (p. 42). *Raw stress* is not very informative, as large values do not necessarily imply a bad model, and is therefore very hard to interpret. To this end, Kruskal (1964) created a normalised measure: *stress-1*. The smaller the stress, the better the fit, with zero stress meaning “perfect fit”. If stress is over 0.2, the model has a “poor” fit (p. 3), which means too much detail is lost in the data reduction.

4.3.1.1 MDS: method

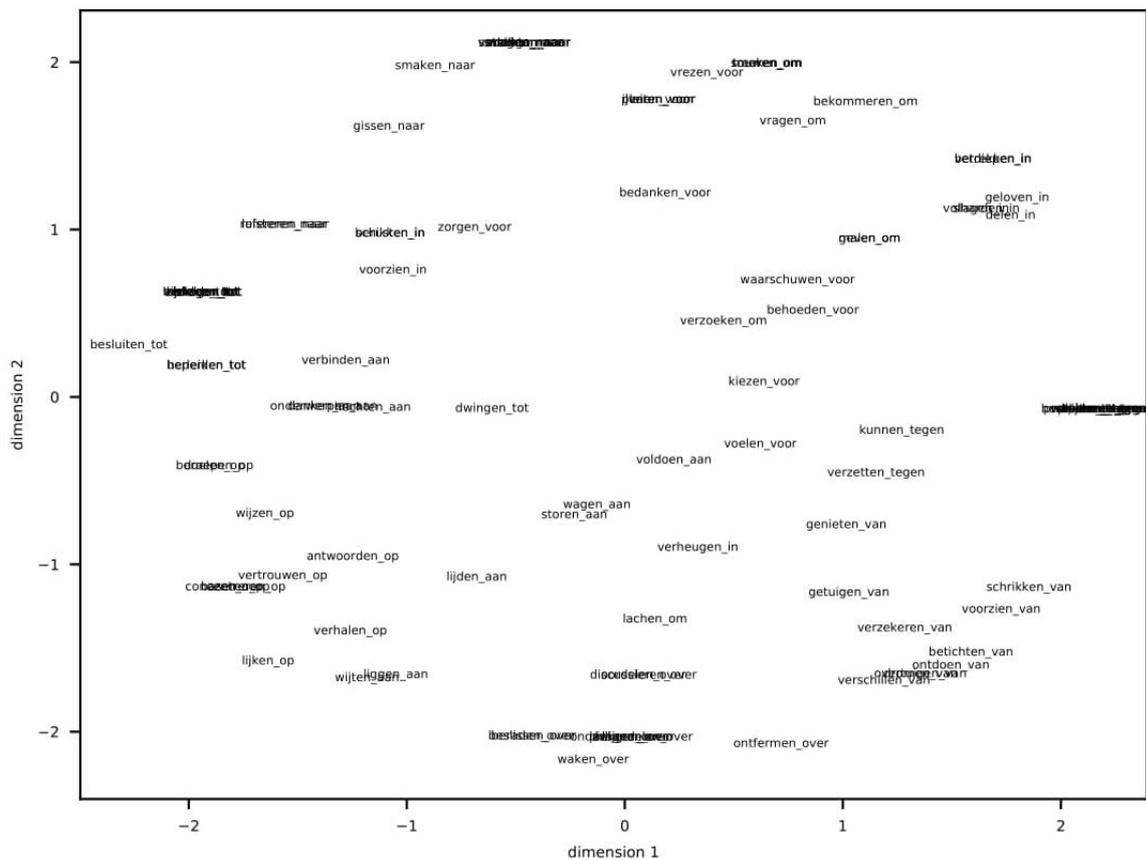
To apply MDS to the current dissimilarity data set, I used the *manifold* module from the machine learning Python library *Scikit-learn* (Pedregosa et al., 2011), which contains a MDS class. It uses a SMACOF (stress majorization) algorithm to transform the matrix into an n -dimensional space. Scikit-learn’s current implementation can only compute and return a value for raw stress, while you would want stress-1, as explained above. In order to be able to assess the fit of the model, I used Borchmann’s (2017) custom version of the class, which allows for the user to set a *normalize* parameter to *True*, and will then return stress-1. The other parameters were set as follows: *dissimilarity* = “precomputed”, *eps* (“Relative tolerance with respect to stress at which to declare convergence.”) = 3.0E-11. The number of “components”—the number of dimensions desired—was varied, as the program was run multiple times. Otherwise, default parameters were used.

4.3.1.2 MDS: Results

First, I ran the MDS program on the dissimilarity data with 2 dimensions. Figure 4.1 presents the scatterplot created from that analysis. The plot does not show clear grouping of predicates, and includes a lot of overlapping ones. Upon further analysis, the stress of this model turned out to be .33, which is over the maximum stress to be considered a suitable model (i.e., .2), which means that the data cannot be reliably visualised in a two-dimensional model. As a three-dimensional model would still be visually interpretable, I ran the program again, with 3 components specified. Unfortunately, stress was still too high (.23). Only when running the program with 4 components did the stress decrease enough for an acceptable—though still not very good—model (.18). However, as a four-dimensional model is not understandable anymore, this defeats the purpose of applying MDS to the data. Therefore, the conclusion must be drawn that, while MDS is a useful tool for other studies, it is simply not the right data visualisation method for this data set. The following sections will describe other ways of interpreting the dissimilarity data.

Figure 4.1

Two-dimensional scatterplot with verb labels from MDS on dissimilarity data.



4.3.2 Colour coded analysis of full dissimilarity matrix

In order to still be able to analyse the data visually, without reading and comparing every cell of a 99×99 matrix, a colour scale was applied to the matrix. The results of this can be found in Figure 4.2. Cells with green colouring represent predicate pairs that are most similar (with a Hamming distance of 0); cells with red colouring are pairs that are very dissimilar (with a Hamming distance of 3). Black lines divide the preposition groups. A few things stand out now. First, as is easily visible, the dissimilarity data forms a clear pattern. Along the main diagonal, cells are greener than above (and thus below)⁷ the diagonal, which means that pairs of predicates with the same preposition have, on average, a lower dissimilarity score than pairs of predicates with different prepositions. Second, the only cells where you see a score of 0, are between pairs of predicates with the same preposition. Last, even though all preposition groups are more similar among themselves, there are a few that are noticeably *more* similar than others—most prominently the group of predicates that combine with *tegen*. Two predicates aside, all pairs in this group have a dissimilarity score of 0, which means they are—with respect to their translations—completely identical.

⁷ As mentioned before, the two halves of the matrix are mirror images of each other.

Figure 4.2

The full dissimilarity matrix with a colour scale applied; green = no dissimilarity, red = completely dissimilar.



To allow for a more quantitative analysis of the data visualised in Figure 4.2, group-level means were calculated for every pair of preposition groups, i.e., the mean of all scores comparing the predicates in preposition group A with the predicates in preposition group B (for example, all cells comparing *aan*-predicates with *in*-predicates). Those means can be found in (the similarly colour coded) Table 4.5.

Table 4.3

The average dissimilarities between all preposition groups; colour marked.

| | <i>aan</i> | <i>in</i> | <i>naar</i> | <i>om</i> | <i>op</i> | <i>over</i> | <i>tegen</i> | <i>tot</i> | <i>van</i> | <i>voor</i> |
|--------------|------------|-----------|-------------|-----------|-----------|-------------|--------------|------------|------------|-------------|
| <i>aan</i> | 1.42 | 2.75 | 2.92 | 2.93 | 2.69 | 2.90 | 2.90 | 2.64 | 2.90 | 2.87 |
| <i>in</i> | 2.75 | 1.29 | 2.93 | 2.91 | 2.89 | 2.89 | 2.96 | 2.75 | 2.95 | 2.93 |
| <i>naar</i> | 2.92 | 2.93 | 1.04 | 2.71 | 2.79 | 2.90 | 3.00 | 2.80 | 2.96 | 2.71 |
| <i>om</i> | 2.93 | 2.91 | 2.71 | 1.26 | 2.98 | 2.77 | 2.94 | 2.99 | 2.93 | 2.66 |
| <i>op</i> | 2.69 | 2.89 | 2.79 | 2.98 | 1.09 | 2.89 | 2.95 | 2.60 | 2.94 | 2.92 |
| <i>over</i> | 2.90 | 2.89 | 2.90 | 2.77 | 2.89 | 0.92 | 2.96 | 2.97 | 2.88 | 2.91 |
| <i>tegen</i> | 2.90 | 2.96 | 3.00 | 2.94 | 2.95 | 2.96 | 0.66 | 2.98 | 2.94 | 2.94 |
| <i>tot</i> | 2.64 | 2.75 | 2.80 | 2.99 | 2.60 | 2.97 | 2.98 | 0.66 | 2.98 | 2.90 |
| <i>van</i> | 2.90 | 2.95 | 2.96 | 2.93 | 2.94 | 2.88 | 2.94 | 2.98 | 1.45 | 2.86 |
| <i>voor</i> | 2.87 | 2.93 | 2.71 | 2.66 | 2.92 | 2.91 | 2.94 | 2.90 | 2.86 | 1.38 |

These means further confirm the first and third observations described above. Dissimilarity between predicates with the same preposition (i.e., intra preposition group dissimilarity) is, in all cases, lower than dissimilarity between predicates with a different preposition (i.e., inter preposition group dissimilarity). The mean dissimilarity score of the former set of groups is 1.12, while the mean of the latter groups is 2.88, which is nearly completely dissimilar. There is some variation in the intra preposition pairs, though. On the one hand, there are some groups with almost no variation at all (*tegen* and *tot*), while, on the other hand, some groups have quite a lot of variation (*aan* and *van*).

Table 4.6 and Table 4.7 display the minimum and maximum values found in each group. As already mentioned above, no 0 dissimilarity values are found outside the main diagonal, but all group pairs along the diagonal contain at least one predicate pair with a value of 0. Other than that, one cell stands out in the minima table: the predicate group pair *naar-tegen*. All predicates in either group are completely different from the ones in the other group. This is actually quite understandable, when you realise that the group with *tegen* was one of the two with the lowest mean dissimilarity. This means that almost all translated predicates combined with the same preposition, and therefore, even if all translated prepositions were represented in the *naar* group (which is not the case), there is less opportunity for overlapping prepositions. In fact, all combinations of the *tegen* group with other groups—other than the comparison with itself, of course—yielded mean dissimilarities of at least 2.90 (all higher than the average dissimilarity score outside of intra preposition group comparisons). The predicates that prevent *tegen* from having a 3 dissimilarity with all other groups, are *kunnen tegen* ‘can stand’ and *verzetten tegen* ‘resist’; with these, Dutch seems to be the outlier in making it a prepositional verb. Both the English and German equivalents (*vertragen* and *widersetzen*, respectively) take a non-prepositional object.

Table 4.4

The minimum dissimilarities between all preposition groups; colour marked.

| | <i>aan</i> | <i>in</i> | <i>naar</i> | <i>om</i> | <i>op</i> | <i>over</i> | <i>tegen</i> | <i>tot</i> | <i>van</i> | <i>Voor</i> |
|--------------|------------|-----------|-------------|-----------|-----------|-------------|--------------|------------|------------|-------------|
| <i>aan</i> | 0.0 | 1.0 | 2.0 | 2.0 | 1.0 | 2.0 | 1.0 | 2.0 | 1.0 | 1.0 |
| <i>in</i> | 1.0 | 0.0 | 2.0 | 2.0 | 2.0 | 1.0 | 2.0 | 2.0 | 2.0 | 2.0 |
| <i>naar</i> | 2.0 | 2.0 | 0.0 | 1.0 | 1.0 | 2.0 | 3.0 | 1.0 | 2.0 | 2.0 |
| <i>om</i> | 2.0 | 2.0 | 1.0 | 0.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 1.0 |
| <i>op</i> | 1.0 | 2.0 | 1.0 | 2.0 | 0.0 | 2.0 | 2.0 | 1.0 | 2.0 | 2.0 |
| <i>over</i> | 2.0 | 1.0 | 2.0 | 2.0 | 2.0 | 0.0 | 2.0 | 2.0 | 1.0 | 2.0 |
| <i>tegen</i> | 1.0 | 2.0 | 3.0 | 2.0 | 2.0 | 2.0 | 0.0 | 2.0 | 1.0 | 1.0 |
| <i>tot</i> | 2.0 | 2.0 | 1.0 | 2.0 | 1.0 | 2.0 | 2.0 | 0.0 | 2.0 | 2.0 |
| <i>van</i> | 1.0 | 2.0 | 2.0 | 2.0 | 2.0 | 1.0 | 1.0 | 2.0 | 0.0 | 1.0 |
| <i>voor</i> | 1.0 | 2.0 | 2.0 | 1.0 | 2.0 | 2.0 | 1.0 | 2.0 | 1.0 | 0.0 |

The maxima table holds no surprises. All predicate group pairs contain at least one predicate pair that is completely dissimilar—with the exception of intra preposition pairs, which, by definition, have the Dutch preposition in common.

Table 4.5

The maximum dissimilarities between all preposition groups; colour marked.

| | <i>aan</i> | <i>in</i> | <i>naar</i> | <i>om</i> | <i>op</i> | <i>over</i> | <i>tegen</i> | <i>tot</i> | <i>van</i> | <i>voor</i> |
|--------------|------------|-----------|-------------|-----------|-----------|-------------|--------------|------------|------------|-------------|
| <i>aan</i> | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| <i>in</i> | 3.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| <i>naar</i> | 3.0 | 3.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| <i>om</i> | 3.0 | 3.0 | 3.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| <i>op</i> | 3.0 | 3.0 | 3.0 | 3.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| <i>over</i> | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| <i>tegen</i> | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 2.0 | 3.0 | 3.0 | 3.0 |
| <i>tot</i> | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 2.0 | 3.0 | 3.0 |
| <i>van</i> | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 2.0 | 3.0 |
| <i>voor</i> | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 2.0 |

4.3.3 Translated prepositions

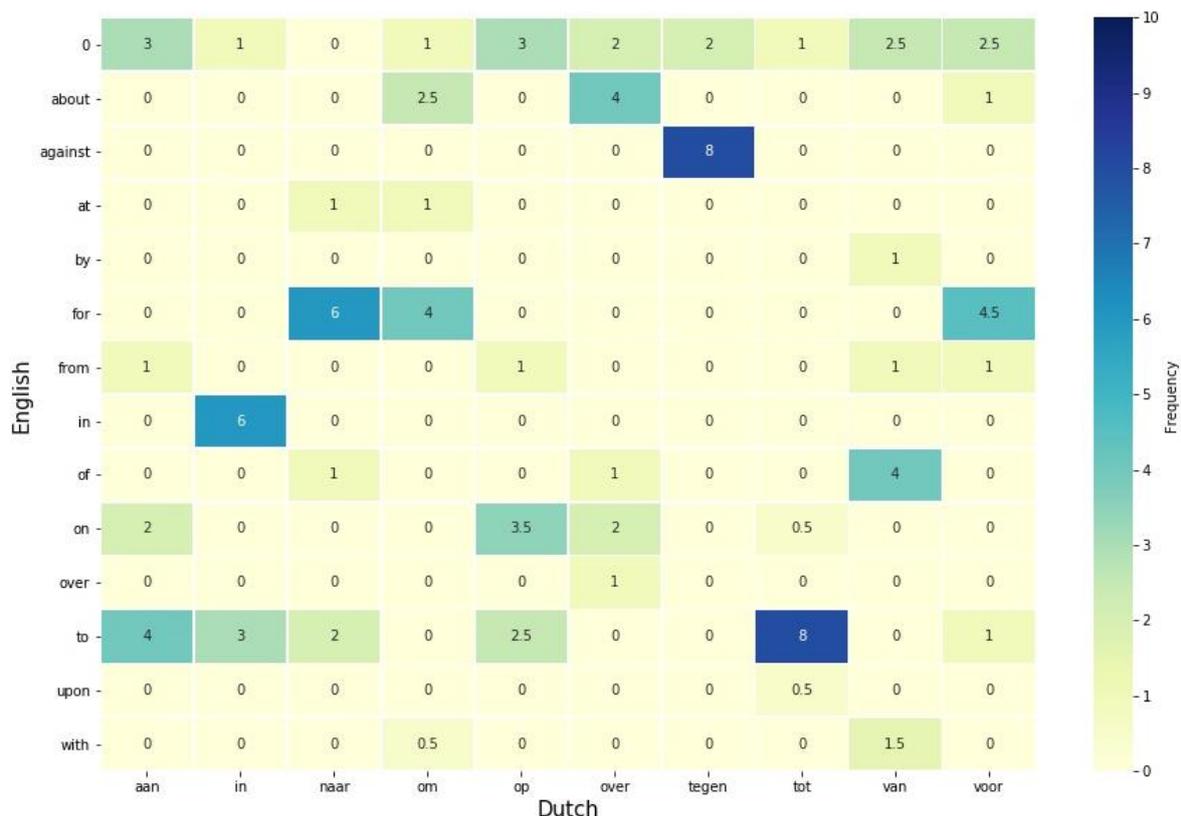
There is one aspect of the cross-linguistic data yet to be discussed: the actual translations. The previous section already showed some interesting patterns in the data; this section will further investigate which prepositions were actually used in the translations.

I ran a frequency analysis on the translated data, the results of which are presented in heatmaps in Figure 4.3 (Dutch to English) and Figure 4.4 (Dutch to German). On the x-axis, these heatmaps list all Dutch preposition, and the y-axis list all translated prepositions represented in this data set. The numbers indicate how often a Dutch preposition is translated with a particular English or German preposition. The darker blue a cell is in this heatmap, the more frequent a certain translation is. For each of the figures, I will discuss some points that stand out.

First, for the Dutch to English translations, you can see that both groups that had the lowest intra preposition group dissimilarity (*tegen* and *tot*) most often have the same translation (*against* and *to*, respectively). For only two of the predicates in these groups, another translation is used. Two other groups that have high-frequency translations (i.e., at least six of the translations use the same preposition) are the groups of predicates with *naar*, which is most often translated with *for* (6 times), and *in* (6 times), which is most often translated with *in*. Even for non Dutch speakers, that last group must serve as a clear example of the second observation we can make: for most high-frequency translations, the preposition is the translational equivalent of the Dutch original.

Figure 4.3

Heatmap of the frequency of the translations from Dutch to English.



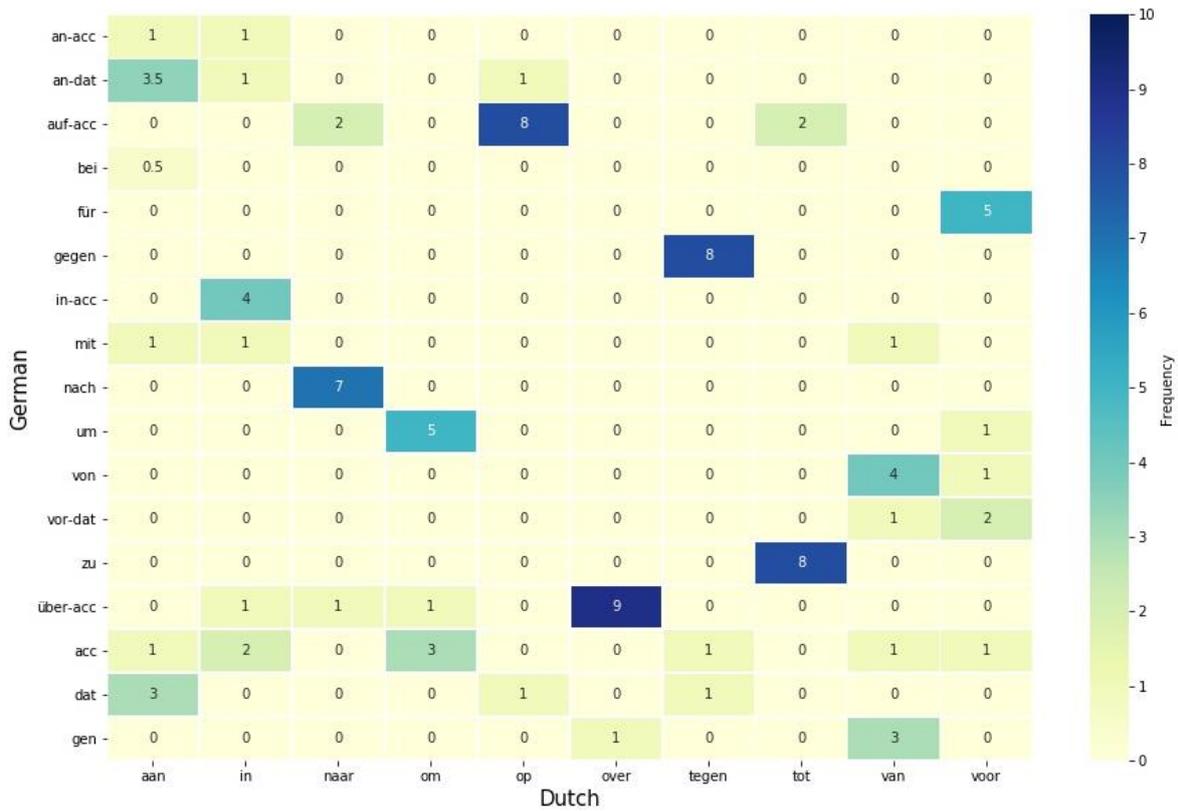
With respect to the German translations, you can see a few more darker blue cells. The groups of predicates with *tegen* and *tot* are, again, among the highest-frequency groups, with both having a translation with a frequency of 8; they are translated most often with *gegen* and *zu*, respectively. There is one group here, however, with an even higher frequency of 9, which is

the group with *over* (translated with *über-acc*). Due to their consistent translations into German, the predicates in this group closely follow predicates with *tegen* and *tot* as the group with the third-lowest intra preposition group dissimilarity. The final two groups with high-frequency translations are *naar* (translated 7 times with *nach*) and *op* (translated 8 times with *auf-acc*). These groups are number four and five, after the three other groups discussed here, in the ranked list of intra preposition group dissimilarity.

You can see that although there are two-way prepositions included in the data, only one of those actually alternates between the dative and accusative case (*an*). All other two-way prepositions only use one of them in all translations. There is, however, variation among those predicates that use a non-prepositional complement, as case marking varies between predicates in the same (Dutch) preposition group. For example, while *wijten* ‘blame’ and *voldoen* ‘meet (e.g., a requirement)’ both govern the preposition *aan*, the former governs a dative object in German (*zuschreiben* + *dat*) and the latter an accusative one (*erfüllen* + *acc*).

Figure 4.4

Heatmap of the frequency of the translations from Dutch to German.



4.4. Discussion

This chapter described a cross-linguistic study on the selected 99 predicates in this thesis. English and German translations were found for each of those predicates, with the help of several dictionaries. A dissimilarity matrix was constructed, which contained a dissimilarity score for each pair of predicates (ranging from 0, completely similar, to 3, completely dissimilar). Multidimensional scaling was applied to this data as a way of visualising the

matrix, but to no avail, as the stress indicated that the model did not fit the data adequately enough.

Instead, a visual analysis with colour coding was done. The results of this showed a clear tendency for predicates with the same preposition to be more similar to each other than to predicates with a different preposition. This means that, overall, when Dutch predicates use the same preposition, English and German do the same, confirming my hypothesis. This is clear evidence towards the idea proposed in this thesis, namely that the process of selecting a preposition is not random.

Furthermore, a closer look at the translations used showed that high-frequency translations are translations equivalents (i.e., occupying a similar semantic space in their respective languages) of the original Dutch preposition (as for instance with *tegen*, *against*, and *gegen*). The conclusion can therefore be drawn that, not only are prepositions not selected at random, their semantics are involved in this process. After all, if semantics of a prepositions were not important, a random preposition would be selected, which does not happen here.

The same translation data also shows that, although English and German both resemble Dutch use of prepositions, German is more similar to Dutch, as there are more and higher high-frequency translations. For English, there are 2 groups of predicates that use the same translations in 8 cases, and 2 groups that use the same preposition in 6 cases. On the other hand, in German, there is one group that uses the same preposition in 9 cases, 3 that do so in 8 cases, and 1 that does so in 7 cases. This closer similarity in prepositional object marking is not unexpected, as Dutch and German are lexically more closely related than Dutch and English. Van Haeringen (1956: 74) mentions that the old Germanic lexicon is largely shared between Dutch and German, while those Germanic elements changed to a much greater extent for English. Additionally, formation of new words happens in largely the same way for Dutch and German, while English makes more use of French and Latin influences in forming new words.

One question that remains is: Why is there such a difference between the intra preposition group dissimilarity score of the different groups? On the lower end you have groups like *tegen* and *tot*, both of which have a dissimilarity score of .66. On the higher end you have groups like *aan* and *van*, with dissimilarity scores of 1.42 and 1.45, respectively. A possible reason is that this is due to the more narrow meaning of the former kind of prepositions, and the broad meaning of the latter.

Loonen (2003, chapter 4.3.1) discusses the different senses of various prepositions in fixed expressions. For *aan*, he lists 15 different senses, such as ‘expressing the cause of some negative effect on the subject’ (*lijden aan* ‘suffer from’, *te gronde gaan aan* ‘be ruined by’) (4.3.1.1, sense 1). Voor *van*, he lists even more senses, totalling to 24, such as ‘expressing the (desire/need for) a separation between subject/object and the complement of the preposition’ (*afzien van* ‘relinquish/waive’, *verlossen van* ‘rid (someone) of’). A preposition like *tegen*, on the other hand, is defined by only 6 senses.

The entries for *aan*, *van*, *tegen*, and *tot* were looked up in the Dutch dictionary *Dikke van Dale* (ref.) and the Van Dale translation dictionaries (Dutch to English and Dutch to German).

Tegen and *tot* list 12 and 10 meanings respectively. The former consists of meanings such as ‘(reverse) direction (whether physical or of expressions)’, ‘opposition’, ‘near something (physical location, moment in time)’, while the latter consists mostly of several

descriptions of ‘goals’, as well as ‘orientation’, and ‘limit’.⁸ Additionally, these prepositions have a clear direct translation. *Tegen* translates to English *against* in 8 out of 9 meanings listed,⁹ and to German *gegen* in 10 out of 10 meanings¹⁰. *Tot* translates to English *to* in 6 out of 7 meanings listed. The German translation is divided between *bis* for the first 3 meanings, and *zu* for the last 4.

The entry in the *Dikke van Dale* for *van*, on the other hand, lists 55 different meanings, including (but, again, not limited to) ‘source/origin’, ‘distance’, ‘descent’, ‘cause’, possession, and ‘means’. The translated prepositions do have a clear ‘winner’ (*of* for English, listed for 12 out of 15 meanings, and *von* for German, listed for all 14 meanings), but I believe that the broad meaning of the preposition allows for its use in such diverse types of predicates, that that influences its cohesiveness. The *Dikke van Dale* entry for *aan* only lists 8 meanings, but here the English translations are different for every meaning listed, resulting in a relatively un-cohesive set of translations.

This illustrates how the English and German equivalents of *tegen* and *tot* occupy a very similar semantic space, while there is no one preposition that covers the same semantic space as *van* or *aan*. It then follows that there would be more variation in the translations of predicates containing one of the latter two prepositions than in translations containing of the former two.

To conclude, this cross-linguistic study provided evidence for the idea that prepositions are not chosen at random, but rather their semantics are at play. The next chapter will describe a diachronic study about a number of the predicates in this thesis to investigate how they developed over time.

⁸ This is not an exhaustive list of all possible meanings of the prepositions. Rather, these are the main points of nuanced meanings combined. For a complete list, please check the (website of) *Dikke van Dale*.

⁹ The number of meanings listed differs from the *Dikke van Dale* to the *Van Dale* translation dictionaries.

¹⁰ This does not mean no other prepositions are mentioned, but all 10 meanings list *gegen* as one of the options, if not the only one.

Chapter 5. Diachronic analysis of prepositional predicates

5.1. Introduction

As Vandeweghe (2015) explains, diachronic data can be used to give context to synchronic data. Chapter 1 of this thesis already discussed Höllein's (2019) finding that prepositional object marking is highly dynamic and an investigation of the topic could benefit from diachronic data. In this chapter, I will supplement the findings of the previous chapters with historical data with the aim to provide a clearer picture of prepositional predicates.

It is generally agreed upon that the loss of case led to a need for a new construction, and prepositions filled this void (Duinhoven, 1989; Hogenhout-Mulder, 1983; Stoet, 1977). Van der Horst (2008) suggests that cases originally expressed a meaning that was later adopted by a preposition, as for instance with the shift from accusative to prepositions that indicated direction, and from the dative, in its instrumental use, to phrases with *met* (p. 461). Vandeweghe (2011) similarly describes how some prepositions are semantically related to the meanings of former Dutch cases, as for example *van* can be connected to the genitive (p. 11).

However, the ways in which prepositional objects have developed differ. Van der Horst (2008) compares *verwonderen* 'marvel', *kijken* 'look/watch' and *zich schamen* 'be embarrassed'. The object of the first has, in all likelihood, developed from a genitive phrase to a prepositional phrase. *Kijken*, on the other hand, was first combined with an adverbial phrase, that later turned into a prepositional object with *naar*. Finally, *zich schamen* used to govern a genitive object, which became an accusative object, before turning into the prepositional phrase *zich om/over iets schamen* in the 18th century, and finally to *zich voor iets schamen* in the 19th century.

This chapter aims to provide an overview of the development of case marking and use of prepositions in general and of the introduction of prepositional predicates. Additionally, a number of predicates will be selected to investigate further using data from several dictionaries. Section 5.2 will discuss how prepositional predicates have developed over time, after which a number of predicates will be selected for a diachronic investigation. Section 5.3 presents which predicates were selected, how these predicates were selected, and what data was gathered. Section 5.4 will present the results from this investigation, and section 5.5 will discuss the results.

5.2. The history of Dutch case marking and prepositions (Van der Horst, 2008)

Van der Horst (2008) extensively describes the evolution of case marking, prepositions, and prepositional objects specifically, in his book on the history of Dutch syntax. His overview starts with Old Dutch (ca. 6th to 12th century), and makes its way, almost century by century, to the 1900s. This section will present the information from his book relevant to this thesis.

Old Dutch used four different cases: nominative, genitive, dative and accusative. However, this was a system already in decline (Van der Horst, 2008: 145), as it had developed from a language which used, in addition to these four, the ablative, vocative, instrumental and locative cases. As Van der Horst describes, Old Dutch also used quite a large number prepositions already, in comparison to other Indo-European languages (he lists 37 prepositions during this time period). A comparison between the *Wachtendonkse Psalmen* (Latin psalms translated to Old Dutch in what is suggested to be the 10th century) and the

Latin original reveals that the translator has, on many occasions, used a preposition where the original text had case marking. You cannot yet speak of a construction like our modern prepositional predicates, but there is one verb that, on occasion, already combines with its ‘own’ preposition: *geloven aan* ‘believe in’, as exemplified in sentence (24).¹¹

- (24) *Nu saget mir einen kuning othir greuen. the aneuren got wille **geloufen**.*
 (Mfr.Reimb. A, r. 358, 1151-1200)
 “Noem mij nu een koning of graaf, die **in jullie God** zou willen **geloven**” [Now name me a king or count, who would want to **believe in your God**]

In Early Middle Dutch, there is a further increase in the use of prepositions.¹² Their number grew, and more and more functions that were previously expressed with case marking are now expressed with prepositions. Furthermore, Van der Horst describes how a shift can be observed in the rise of a “second syntactic valence”. Initially, PPs were independent adverbial clauses, but after some time, PPs can also be found as attributives to nouns. He interprets this development as an increase in valency. This was not yet possible in Old Dutch, but started with *van* in Early Middle Dutch (which was used besides a genitive construction), and was used on larger scale in Middle Dutch (14th and 15th century). In a similar way did prepositional predicates come to exist. Van der Horst (p. 462) wrote: “[...] naast de band met het erop volgende substantief kan het voorzetsel meer en meer ook een band aangaan met een tweede element (substantief, adjectief of werkwoord)” [next to its connection to the following noun, the preposition can, increasingly, form a bond with a second element (noun, adjective, or verb [emphasis added])]. In Early Middle Dutch, there are a handful of more or less fixed verb-preposition combinations (Van der Horst names, in addition to *geloven aan*, *haken naar* ‘long for’). However, Van der Horst also explains that virtually all prepositional predicates we know today did not exist yet in these periods, or were constructed in a different manner.

As previously stated, the preposition *van* is used alongside the genitive in Middle Dutch. For Early Middle Dutch, Van der Horst argues that it is reasonable to assume that the two constructions differ in meaning, but that seems to be less true in Middle Dutch. Compare, for example, the late 15th century sentences (25) and (26), which were found in the same text by a single author; the former sentence uses a prepositional construction while the latter uses the genitive for what is essentially the same phrase.

- (25) *dat zou [zij] van haren rechte bilics [...] **ghebruken mach*** (GB 204,354)
 “Dat zij naar recht en billijkheid **van haar recht gebruik kan maken**” [That she can **exercise her right** in a fair and just manner]
- (26) *[zij] moghen vrijlic haers rechts **ghebruken*** (GB 204,368)
 “[Zij] kunnen ongehinderd **van hun recht gebruik maken**” [[They] can **exercise their right** unimpededly]

The rise of the category of prepositional predicates slowly picks up its pace in late Middle Dutch. New combinations are found, such as *houden van* ‘love’ and *denken aan* ‘think of’, although there are also verbs that vary quite a lot in how their object is marked. For instance,

¹¹ In modern Dutch, this meaning is expressed with a combination with the preposition *in*.

¹² Early Middle Dutch is usually defined as the period 1200 to 1300, but is here used as a name for the first period of Middle Dutch Van der Horst describes (1200 to 1350).

hoeden ‘protect’ can govern a genitive object, a prepositional object with *van*, or a prepositional object with *voor*. You can also regularly observe verb-preposition combinations that were either random, or existed for a time but then gave way to other constructions (e.g., *denken na* ‘think about’, *hopen aan* ‘hope for’, *wensen na* ‘wish for’).

The 16th century brings a further decline of case marking. It has almost disappeared from spoken Dutch (with the exception of certain almost idiomatic expressions). The group of prepositional predicates, however, grows steadily, with predicates like *letten op* ‘pay attention to’, *luisteren naar* ‘listen to’, *spotten met* ‘mock’. However, this ‘fixedness’ of prepositions is not the same for everyone, as the specific preposition used varies sometimes between authors.

This variation carries through to the 17th century, where a vast increase of the number of prepositional predicates can be observed. These predicates are used increasingly often, but one can regularly see a verb accompanied by a preposition other than the usual combination. Van der Horst argues this speaks to the productive use of creating verb-preposition combinations, as people were actively creating new such combinations. This general tendency to create verb-preposition combinations can also be seen in cases in which a preposition was added to a verb that does not govern a preposition today. Also, many verbs that nowadays exclusively use one preposition, went through a phase in which it was used with two, three or even four different prepositions. For example, *wachten* ‘wait’ is used with both *op* and *naar*.

Where data from the 17th century already showed loss of case marking in informal speech, there is further deflection in the 18th century. For this reason, the group of verbs with a fixed preposition instead of a case marked object has grown quite large (although a considerable number of these govern a different preposition than they do presently). There is also, however, a (small) number of verbs that sometimes still used case marking. Van der Horst (p. 1501) gives the example of Huydecoper’s criticism (1730, p. 260) of sentence (27), in which Huydecoper would have preferred the object to be marked with dative case instead of accusative—but does not entertain the idea of a prepositional object.

- (27) *Wy hoeven zeker ons dien schoonzoön niet te **schaemen**.* (Vondel)
“Wij hoeven ons zeker niet **voor** onze schoonzoön te **schamen**.” [We certainly don’t have to **be ashamed of** our son-in-law.]

At the end of the 19th century and first half of the 20th century, case marking had long only been used in written language. Even there, however, it is practically no longer productive: only two forms can be distinguished (and really only with masculine words): those with an *-n* and those without it. On the genitive and feminine dative forms with *-r* that Bilderdijk (1826) lists, he remarks that they are very formal, and would not be used in normal conversation (p. 92-95). Over the course of the 20th century, case marking was finally abolished (first partly in the prescriptive *Spelling-Marchant* of 1934 and then fully in the *Wet voorschriften schrijfwijze Nederlandsche taal* ‘Dutch spelling regulations act’ of 1947). The case marking system had developed into one with three nominal functions (subject, direct and indirect object), and, finally, to a system with two nominal functions (subject and object).

During this period, the number of prepositions available has grown enormously, from the 37 in Old Dutch to 97 in the 20th century. There are a lot of verb-preposition combinations that are the same as today, although some variation still exists in whether or not a preposition is used, or which preposition is used (in some cases, this variation is regional). There are a handful of prepositional predicates that have lost their preposition over time (and not gained another): *raadplegen* (*met, bij*) ‘consult’, *indenken* (*in*) ‘imagine’, and *trachten* (*naar*) ‘try’.

Although the overview by Van der Horst stops there, the category has not stopped developing. Duinhoven (1989) still describes it as a “zinspatroon in wording” [sentence pattern in the making]. The following sections describe a closer historical investigation of a few of the predicates used in this thesis.

5.3. Method

5.3.1 Verb selection

The cross-linguistic study revealed that there was quite some difference between predicates in different preposition groups: some groups were very cohesive (like the group of predicates that govern *tegen*) and in some groups there was a lot of variation. For this diachronic investigation, I selected both predicates that were very similar to other predicates that governed the same preposition, and predicates that were very different from other predicates in the same group.

Remember that the dissimilarity scores were given to a pair of predicates, not to an individual predicate. In order to identify the predicates that were most and least similar to others in their group, a new score needed to be calculated. Therefore, based on the predicate pair dissimilarity scores from the cross-linguistic study, a single *intra preposition group dissimilarity score* was calculated for each predicate. This score was the mean of the of all scores that compared a given predicate to others in its preposition group. See, for example, Figure 5.1. All coloured pixels represent the dissimilarity score of a predicate pair (where dark green is 0, completely similar, red would be 3, completely dissimilar, and light green, yellow, and orange represent the points between: 1, 1.5, and 2, respectively). The first predicate in the set is *danken aan*, which can be found in both the first row (marked with a red box) and the first column of the *aan* predicates. If you want to calculate the mean dissimilarity score of *danken aan* with other *aan* predicates, you need only add up all the scores in either the first row or column (as these are identical) and divide by the number of predicates in the *aan* group. For *danken aan*, this score is 1.30. An overview of all intra preposition group dissimilarity scores can be found in Appendix E.

Figure 5.1

The intra preposition group dissimilarity score of the predicate danken aan (in the first row) is the mean of all the dissimilarity scores in the red box (all those that compare danken aan to other predicates in the aan group).



For the diachronic study in this chapter, I selected four predicates from three groups each: the group with the highest coherence in the cross-linguistic data (*tegen*; mean dissimilarity = 0.66), the group with the lowest coherence (*van*, mean dissimilarity = 1.45), and the group with an average coherence (*op*, mean dissimilarity = 1.09). From each group, the two predicates with the lowest intra preposition group dissimilarity (i.e., those that were most similar to other predicates from the same preposition group) were selected, and the two predicates with the highest intra preposition group dissimilarity (i.e., those that were least similar to other predicates from the same preposition group).¹³ This resulted in a selection of the following 12 predicates: *azen op*, *lijken op*, *verhalen op*, *vertrouwen op*, *beschermen tegen*, *kunnen tegen*, *zich verweren tegen*, *verzetten tegen*, *dromen van*, *overtuigen van*, *schrikken van*, and *voorzien van*. Their intra preposition group dissimilarities are presented in Table 5.1.

Table 5.1

The intra preposition group dissimilarities of the twelve predicates under investigation.

| Preposition | Verb | Intra-P dissimilarity |
|-------------|---------------|-----------------------|
| Op | Azen | 0.85 |
| | Lijken | 1.80 |
| | Verhalen | 1.80 |
| | Vertrouwen | 0.88 |
| Tegen | Beschermen | 0.40 |
| | Kunnen | 1.70 |
| | Zich verweren | 0.40 |
| | Verzetten | 1.70 |
| Van | Dromen | 1.20 |
| | Overtuigen | 1.20 |
| | Schrikken | 1.80 |
| | Voorzien | 1.75 |

5.3.2 Data

I searched the *Oudnederlands Woordenboek* (ONW; Old Dutch Dictionary), *Vroegmiddelnederlands Woordenboek* (VMNW, Early Middle Dutch Dictionary), *Middelnederlands Woordenboek* (MNW, Middle Dutch Dictionary), and *Woordenboek der Nederlandsche Taal* (WNT, Dutch Dictionary), using the website <http://gtb.inl.nl/search/>.¹⁴

Not all predicates were represented in all dictionaries. In fact, the ONW only contains (relevant) data for one of these predicates (i.e., *beschermen*). Secondly, sometimes the VMNW contains more data on a given predicate than the MNW, or it contains data where the latter doesn't. While that might seem counter-intuitive, as one would expect to find more data in the dictionary on more recent language, it can be easily explained. The VMNW is based on all material found from the time period in question (13th century), so it is based on a lot of data. The MNW, on the other hand, is a dictionary written two centuries ago, for which exemplar

¹³ With the exception of *vertrouwen op* (mean dissimilarity = 0.88), which was selected over *zich baseren op* (mean dissimilarity = 0.85) and *zich concentreren op* (mean dissimilarity = 0.85), as only relatively recent language data could be found for the latter two.

¹⁴ All dictionaries were consulted in November 2020.

sentences were collected. Therefore, it contains fewer example sentences, but more importantly, is based on the editor's intuition rather than on all data available. The quantity, quality, and even presence of data in it is therefore highly dependent on the editor working on that subsection of the dictionary. Meaning of words, or even words themselves, are more easily overlooked, and therefore missing in the dictionary. Data from the WNT faces the same challenges.

5.3.3 Analysis

For each predicate, the intended meaning was looked up in the dictionary. For some predicates, many other meanings were also listed—as for example with *kunnen*, which is also a modal verb when not used with the preposition *tegen*—but these are not included in the data presented here. This also means that, sometimes, a predicate can also occur in other constructions than mentioned here, but those constructions have a different meaning and are otherwise not related to the constructions discussed.

In some cases, the dictionaries list the constructions in which a given predicate can be found, but most of the time the data presented is based on the example sentences given. All possible constructions are presented, as well as the first (and sometimes last) attestation of those constructions, together with information on the evolution of the predicate. Section 5.4 will present this data for each of the twelve predicates.

5.4. Results

5.4.1 *Azen op* 'have one's eye on'

The earliest attestation of *azen* in these dictionaries is the sentence in (28) from 1287. It is used here with an object phrase introduced by *naar*.¹⁵ This is the only sentence with this construction, and is it no longer mentioned in any of the more recent dictionaries. After a gap of data in the MNW, data from the WNT shows that a construction with *op*, as in (29), is now the only one with the intended meaning. The earliest attestation of this construction comes from 1727, and no other construction is used after. The meaning of the predicate *azen op* later evolved from the very literal *prey* to the metaphorical use *have one's eye on*, as in (30), but that did not involve a change in preposition.

- (28) *experimentator die seghet dat die crane te roepene pleghet als i .i. nat swerc siet effen hi wil dat dandre alle beseffen ende si **hasen na hare eten** ende als i omme ardinge beeten hare leedere staet inder hoede (... dan staat hun leider op de uitkijk.).* (Der Naturen Bloeme, Maerlant, 1287, p. 173, r. 8-14)
“Experimentator die zegt dat de kraanvogel gewoon is te roepen als hij een regenwolk ziet oprijzen. Hij wil dat de anderen dat bemerken als zij **azen op** (uitgaan om) hun voedsel. En als zij omwille van het voedsel neerstrijken dan staat hun leider op de uitkijk.” [Experimenter who says that the crane is used to

¹⁵ It should be noted that *naar* was not spelled ‘naar’ in these times, but rather ‘na’ or ‘naer’. For a lot of the prepositions discussed in this section, there were multiple spelling variations (e.g., ‘aan’ vs. ‘aen’ vs. ‘an’, and ‘op’ vs. ‘up’), or the modern version was not yet used altogether. However, for the sake of consistency, I will use the modern spelling of all prepositions (examples contain the spelling of the original text).

calling when he sees a rain cloud rising. He wants the others to notice that when they **hunt for** (go out for) their food. And when they settle down for the sake of food, their leader is on the lookout.]

- (29) *Een doode Walvisch ... , waar op een menigte Meeuwen aasden, [...].* (Groenl. Vissch. 139, 1727)
 “A dead whale ..., **on which** a group of seagulls **preyed**, [...].”
- (30) *De vorsten ... , die azen op ons bloed, [...].* (tollens 8, 21 [1839]).
 “The monarchs ..., who **are after** our blood, [...].”

5.4.2 Lijken op ‘resemble’

The oldest sentence in the dictionaries that contain *lijken* accompanied by an object phrase is from 1475~1485 (in (31)). This is also the only instance found where it combines with *tegen*. The dictionaries describe that this verb first used a dative to indicate the element which the subject resembles, until around the 17th century. It then lost most case marking, and the verb developed into a copula with the meaning ‘seem’. An example of one of the latest uses of this construction is presented in (32).

- (31) *Wat minne mach daer liken tieghen.* (Hild. 1475-1485, 93, 92)
 “Welke liefde kan daarmee vergeleken worden.” [What love can be **compared to that**.]
- (32) *Sy **lijcken** vast Aesopus Hondt, Die op een smalle Vonder stondt* (Deugdensp. 245 [1645])
 “Zij **lijken** sterk **op de hond van Aesopus** die op een smal, los bruggetje stond.” [They closely **resemble** Aesopus’ dog who was standing on a small, loose bridge.]

Further, a prepositional phrase with *naar* is used quite a lot in the 18th and 19th century (as in (33)), and finally, the construction *lijken op*, as we know it now, is first found in 1840 (presented in (34)). Two sentences are also mentioned that contain the combination *lijken als*, but the WNT quickly identifies those as a contamination between *lijken op* and *eruitzien als* ‘look like’. The construction with *naar* in (33) also resembles *eruitzien*, which can be realised in the form *ernaar uitzien*. This construction has developed from indicating visual similarity—as is also the case *lijken naar* in (33)—to inferential evidentiality from the 19th century onwards (Faber, Rem, & De Hoop, 2022).

- (33) *De appel **lykt** wel **naar een kolksche-koek**.* (Wolff en Deken, 1784, Leev. 1, 195)
 “The apple **resembles** a biscuit.”
- (34) *[Hij] Klom ... op 't groen stellage, 't **Lijkt** krek **op een preekstoel**.* (v. zeggelen 1840, 4, 43)
 “Hij klom op de groene stellage/stelling/steiger. Die **lijkt** sprekend **op een preekstoel**.” [He climbed onto the green construction. It **resembles** a pulpit.]

It seems however, that *lijken* is actually the more modern form of the verb *geli(j)ken*, which lost its prefix *ge-* over time. Data shows that *geli(j)ken* also governed the dative and *naar*. It turns out, however, that the construction (*ge*)*lijken naar* is actually quite a bit older, as (35)

presents a sentence containing the construction from 1287. It existed thus alongside the dative construction, although the data suggests the case marking construction was more common during that period. The MNW, for example, only mentions the prepositional phrase as what seems like an afterthought. In that same period, a combination with *bij* was also possible.

- (35) *siminkele vintmen some ghestart somo (l. some) scone die sijn ghebard met breden starte **ghelikende** clene **na die andre ghemene**.* (Nat.Bl.D., 1287, p. 115, r. 24-27)
 “Sommige apen hebben een staart. Sommige mooie [apen] hebben een baard en een brede staart. Zij **lijken** niet **op de vorige** (eerst genoemde apen)”. [Some monkeys have a tail. Some beautiful [monkeys] have a beard and a broad tail. They do not **resemble the previous**[ly mentioned monkeys].]

The construction *gelijken op* also turns out to be a bit older than *lijken op*, as its first attestation lies in 1785. This sentence is shown in (36). It seems like the verb only lost its prefix after the combination with *op* became common.

- (36) *Hoe weinig hy ook **op zyne zuster gelykt**, nu zag ik, en dat al vry sterk, de familietrek!* (wolff en deken, 1785, *Leev.* 8, 192)
 “Hoe weinig hij ook **lijkt op zijn zuster**, opeens zag ik toch, en ook vrij sterk, de familietrek!” [As little as he **resembles his sister**, I suddenly saw, and quite strongly, the family trait!]

Finally, the combination *(ge)lijken aan* is also discussed, but this quickly disappeared again. The WNT determines this construction was a Gallicism, and came into existence under the influence of the French *ressembler à* ‘look like’.

5.4.3 *Verhalen op* ‘recover from’

Verhalen occurs quite early in prepositional constructions. It is already found in the 13th century with *op* and *aan*. An early example of *verhalen op*, which is the construction still used today, is given in (37). An example of *verhalen aan* is given in (38). Those constructions existed alongside each other for some centuries. The last example of *verhalen aan* in the dictionaries dates from 1809.

- (37) *So wi dat niet ne dade dat sovden wi **op hem verhalen**.* (Corp.I, 1289, p. 1415, r. 7-8)
 “En als iemand dat niet deed, dan zouden we dat **op hem verhalen**.” [And if someone did not do that, we would **blame him**.]
- (38) *Dis presenteric mine bede V. maget schone ende vtuerkoren Ende dat ic hebbe an v uerloren Laett mi **an v verhalen** doch.* (Lutg.K, 1265-1270, p. 483, r. 22-25)
 “Daarom draag ik een gebed aan u op innerlijk mooie en uitverkoren maagd. Dat wat ik ten op zichte van u heb geruïneerd, laat mij dat herstellen.” [Therefore, I dedicate a prayer to you on inner beautiful and chosen virgin. That which I have ruined before you, let me restore it.]

Over the years, a few other prepositional constructions have appeared and disappeared. For instance, (39) presents a sentence with *tegen*, and (40) with *in*. Other combinations that have been attested are *verhalen* with *van* and *met*. However, in the last few centuries, *verhalen op* had the clear majority.

- (39) *Dat des Sondaeghs inde Halle geene Gilde-Broeders sullen vermogen Vlees te verkoopen, maer dat de Halle gesloten sal moeten blijven den geheelen dag gedurende; op de verbeurte van twaelf Guldens: Welcke boete mijn Heere de Schout ofte sijne Substituten **tegens den Overluyden** ... **verhalen sullen**.* (Handv. v. Amst., 1672, 118 b)
 “Dat op zondag de gildebroeders geen vlees mogen verkopen in de hal en dat de hal de hele dag gesloten moet zijn; anders volgt er een boete van 12 gulden. Deze boete zal mijnheer de schout of zijn vervanger **verhalen op de hoofdmannen** van het gilde.” [That on Sunday the guild brothers may not sell meat in the hall and that the hall must be closed all day; otherwise a fine of 12 guilders will follow. This the master of the guild or his replacement will **recoup** this fine **from the headmen of the guild**.]
- (40) *Zoo 't lot alhier my blijven deed, Ik zoude **in Osmans bloed** die wrange smart **verhalen**. En licht op Djaffirs hoofd uw weig'ring neêr doen dalen.* (v. lennep, 1826, Poët. 6, 140)
 “Als het lot mij hier zou laten blijven, dan zou ik genoegdoening voor de scherpe pijn krijgen uit het bloed van Osman en licht op het hoofd van Djaffirs uw weigering laten neerdalen.” [If fate would let me stay here, I would receive satisfaction for the sharp pain from the blood of Osman and let your refusal fall lightly on the head of Djaffirs.]

5.4.4 *Vertrouwen op* ‘trust/rely on’

The verb *vertrouwen* was formed in the 16th century from *betrouwen* and *getrouwen* (with the same meaning), very likely under the influence of the German *vertrauen* ‘trust’. Both *betrouwen* and *getrouwen* govern a dative object, as in (41). Additionally, *betrouwen* can also take a prepositional object with *op*, which was first attested around 1300 (presented in (42)).

- (41) *Hoet mi, vrouwe, **up u ic mi betrauwe**.* (OVl., 1340-1360, Lied. e. Ged. 54, 119)
 “Bescherm mij, vrouw, ik **vertrouw op u**.” [Protect me, lady, I **rely on you**.]
- (42) *Min harte wille dat ic die cracht An gode soeke van hemmelrike Hie salse gheuen ghe+trouwelijke Als hie al den anderen doet Die **vp hem betrouwen** goed.* (Alexius, 1291-1300, p. 527, r. 10-14)
 “Mijn hart wil dat ik de kracht bij God zoek. Hij zal haar geven vol vertrouwen, zoals hij ook de anderen bedeeft die **op hem vertrouwen**.” [My heart wants me to seek strength from God. He will give it with confidence, as he also gives to others who **trust in him**.]

When *vertrouwen* was formed, it was not used with a dative object, but only with *op*, as in (43), and *in*, as in (44). Both forms still exist today, and similarly to that time period the construction with *op* is considerably more frequent than the one with *in*.

- (43) *Wy selue hadden in onsseluen het vonnisse des doots: op dat wy niet **op onsseluen vertrouwen** en souden, maer op God die de dooden verweckt.* (Bijbel v. Deux Aes, 1562, 2 Cor. 1, 9)
 “Wij hadden in onszelf het vonnis van de dood, zodat wij niet **op onszelf** zouden **vertrouwen**, maar op God die de doden opwekt uit de dood.” [We had within ourselves the sentence of death, so that we should not **trust in ourselves**, but in God who raises the dead from death.]
- (44) *Tfy, schaemt u der sonden, dat gyse opschrijven doet, Die uwe stinkende God Melis **daerin** gy **vertrout**, Niet eeren.* (Geuzenliedb., 1566, 1, 24)
 “Foei, schaam u voor die zonde, dat u hen veroordeelt die uw stinkende God Melis, **waarop** u **vertrouwt**, niet eren.” [Fie, be ashamed of that sin, that you condemn those who do not honor your stinking God Melis, **in whom** you **trust**.]

5.4.5 Beschermen tegen ‘protect against’

Beschermen was the only verb with relevant information in the ONW, so there is rather early data available here. One of the first attestations were sentences that contained the preposition *with* (abbreviated version of *wither*, modern Dutch *weder*, which has a similar meaning as *tegen*) or a Latin preposition (*ab* or *contra*), as some sentences were half Old Dutch, half Latin. Two examples even show the use of *with(er)* and *contra* in the same sentence, as in (45). In this same period, a sentence with *van* was also attested (10th century), which is presented in (46). There are no sentences with case marking, nor is it mentioned in the ONW, VMNW, or MNW. The WNT, however, mentions how *beschermen* used to have a genitive object. This would explain the use of *van* early on, as previous literature has connected the meanings of those two constructions (e.g., Vandeweghe, 2011). This construction with *van* seems to have been the norm for a few centuries, but finally disappeared in the 15th century.

- (45) *Thaz godesbedde bewarent in thirro wereldthimsternisse thie sancti doctores, ande thie sculan erwelet sin ex fortissimis Israhel, wanda thie sculan wesan rectores animarum, thie the cristanheyd cunnan **beskirman** mit spiritualibus armis **with** thene diuvel ande **contra hereticos**.* (LW 051, ca. 1100, 10-11)
 “The holy teachers protect the bed of God in this earthly darkness, and they have to be chosen from the strongest of Israel, for they need to be the teachers of souls, who can **protect** the Christian community with spiritual weapons **from the devil** and **from heretics**.”
- (46) *Thu **beschirmedos** mi **fan** samnungun ueldadigero, **fan** menege uuirkindere unreht.* (WPs (hs. H), 901-1000 (copy 1595-1605) 063,03)
 “You **protected** me **against** the community of malefactors, **against** the crows of wrongdoers.”

From the 13th century onward we also see sentences with *tegen*, as in (47), and from the 14th century also with *voor*, as in (48). These constructions therefore existed at the same time as the one with *van*, but the latter gave way in the end. Interestingly enough, a third sentence was found where *beschermen* governs two different prepositions—in this case *van* and *tegen*—which is presented in (49). Both the construction with *tegen* and with *voor* still exist to this day.

- (47) *bi hem (t.w. een kasteel) was **jegen** die viande **bescermt** ende bereht dat lant.* (Wr.Rag. p. 367, r. 16-17, 1260-1280)
 “door hem was het land **tegen** de vijand **beschermd** en in staat van paraatheid gebracht” [Through him [i.e. a castle] the land was **protected against** the enemy and readied.]
- (48) *Te hem, diene van messciene al **bescermet** voer die honde.* (Nat. Bl. XII 1351-1375, 756)
 “Tot hem [waarschijnlijk een steen], die hen van ongeluk **beschermt** **tegen** de honden.” [To him [likely a rock], who **protects** them from misfortune **against** the dogs.]
- (49) *Dat hy mit synen lyve dusent scepe .. **bescermt** hadde van den brande **teghen** alle onse viande.* (Troyen 7694, 1470-1480)
 “Dat hij met zijn lichaam duizend schepen **beschermd** had **tegen** brand en **tegen** al onze vijanden.” [That he **protected** a thousand ships with his body **against** fire and **against** al our enemies.]

5.4.6 Kunnen tegen ‘can stand’

This construction is the most recent one of this set. It is discussed in dictionaries extensively from quite early on, but not with this meaning. The earliest attestation with *tegen*, presented in (50), dates from 1806. There has, however, never been any doubt about which preposition to use, as no other options are provided.

- (50) *Waarlijk de mensch **kan** ‘er niet **tegen**, dat hij zoo hoog getild wordt.* (Loosjes, Bronkh. 1, 57, 1806)
 “Werkelijk, de mens kan het niet verdragen dat hij zo hoog verheft wordt.” [Truly, man cannot bear to be exalted so highly.]

5.4.7 Verweren tegen ‘defend against’

It seems that *verweren* first used a genitive object, as in (51), often in the form of the word *des*. At the same time, however, the construction *verweren van* and *verweren tegen* also existed, as in (52). This construction held, and is still used today. In the same period, one instance of *verweren voor* is attested, as in (53), but it cannot be found in a later period. The MNW contains very few sentences with an object (many are sentences along the lines of ‘when X happens, someone is permitted to defend themselves’), so it is not clear when the construction with case marking became obsolete.

- (51) *Mar een iosen ende een cares. Begonsten hem **uerwerren** des. Si waren vorbaerst van der stede. Tander volc dat help hem mede.* (Rijmb. p. 706, r. 6-9, 1285)
 “Maar een Josephus en een Chares begonnen zich **daartegen** te **verdedigen**. Zij waren de oudsten (voornaamsten) van de stad. De rest van het volk hielp mee.” [But a Josephus and a Chares started to **defend** themselves **from** that. They were the oldest [most prominent] of the city. The rest of the people helped.]

- (52) *Jnde nũ heuet di hi angst van den drake. Jnde wagt buten ant water. Jnde **verwert** hũme **tegens den drake**.* (Moraalb. p. 420, r. 8-9, Nederrijn, 1270-1290)
 “En nu heeft hij angst voor de draak. En waakt buiten aan het water en **verdedigt** zich **tegen de draak**.” [And now he fears the dragon. And keeps vigil by the water outside and **protects** himself **from the dragon**.]
- (53) *Die scepenen vorseit ne wilden niet ontfaen. so+dat hem die vorseide jan vten sacke moeste **verweren vor ons heren skonings executeurs**.* (Corp.I p. 2249, r. 3-5, 1295)
 “De schepenen die hier voorgenoemd zijn wilden hem niet ontvangen zodat de voorgenoemde Jan Uit de Zak zich moest **verdedigen tegenover de executeurs van de koning**.” [The aforementioned aldermen did not want to receive him, so that the aforementioned “Jan Uit de Zak” had to **defend** himself **against the king’s executors**.]

A final construction of *verweren van* is used sometimes in the 17th and 18th century. This combination might have formed under the influence of the French *se défendre de*, but it gave way to the much more frequent construction with *tegen*.

5.4.8 *Zich verzetten tegen* ‘resist’

Verzetten was formed from the verb *zetten*, after adding the prefix *ver-* for emphasis. The data for both *zetten* and *verzetten* is rather straightforward. Both verbs are only used in combination with *tegen*, already since the 13th century. Sentences (54) and (55) contain two examples of both verbs in this construction. *Verzetten* was found from the 14th century, while *zetten* is attested with this meaning until the end of the 19th century, so they co-existed for quite a few centuries.

- (54) ***Jeghen hem setten** de clerken te Rome, die heren wilden sijn.* (Sp. III 2, 23, 50, 1301-1325)
 “**Tegen hem verzetten** zich de klerken in Rome, die de baas wilden zijn.” [The clerks in Rome, who wanted to rule, **opposed him**.]
- (55) *Niemen dorste vortmee letten noch ooc **jeghen hem versetten**.* (Sp. I 4, 29, 15, 1301-1325)
 “Niemand durfde voorts tegen te werken of zich **tegen hem te verzetten**.” [No one dared to oppose or **resist him**.]

5.4.9 *Dromen van* ‘dream of’

This is another verb of which the meaning evolved from a very literal interpretation to a metaphorical one. The construction *dromen van* already existed in the 13th century, as in (56), but with the literal meaning ‘imagine while sleeping’ that would be expressed with *dromen over* today. From the 17th century the construction gained the metaphorical meaning ‘wish for’, as exemplified in (57). No other constructions are attested in these dictionaries.

- (56) *Hem was te moede no min no mere. Dan als eenen die word onwake. Ende **ghedroemt** heuet **van suare sake**.* (Rijmb. p. 78, r. 24-26, 1285)

“Hij voelde zich precies als iemand die wakker wordt en **over** akelige dingen gedroomd heeft.” [He feels exactly like someone who wakes up and has **dreamed about** afwul things.]

- (57) *Slechts van slemp' en smeeren droomen.* (westerbaen, Ged. 1, 43, 1624)
“Alleen maar **van** slampen en schransen **dromen**.” [Only **dream about** drinking and eating.]

5.4.10 Overtuigen van ‘convince of’

The earliest attestation of *overtuigen* shows a combination with the genitive form *des* in the 14th century. This sentence is presented in (58). After that, it seems that case marking decreased, and unmarked objects were used. In the 17th century, the modern construction with *van* was introduced, as in (59).

- (58) *Were dat enich man vor den richter ... [missing words] ..., des men hem mitten richter ... **overtugen** mach, des en mach hem nyemen onsculdichen.* (Nijh. 1, 228, 1327)
“Als iemand voor de rechter (...) waarvan men hem in tegenwoordigheid van de rechter ... **overtuigen** kan, daarvan (van een aanklacht) kan zich niemand door een eed zuiveren.” [If one is before the judge (...) **of which** he can be **convinced** in the presence of the judge ..., of that (accusation) no one can clear himself by an oath.]
- (59) *Hy was de ledicheit zoo geweldigh toegedaan, dat hy Gods uitspraak ... in 't zweet uws aanzichts zult ghy uw brood winnen, **van** onwaarheid poogde **t'overtuigen**.* (de brune, Jok en E. 210, 1644)
“Hij koesterde het nietsdoen in zulke hoge mate dat hij probeerde te bewijzen/overtuigen dat de uitspraak van God ‘In het zweet des aanschijns zult gij uw brood verdienen’, onwaar was.” [He cherished idleness to such an extent that he tried to prove/convince that the statement of God “In the sweat of your brow you shall earn your living” was untrue.]

5.4.11 Schrikken van ‘get scared by’

The earliest attestations of *schrikken* use combinations with *van* and *uit*. However, these are again very literal, in that they expressed that someone jerked away from something, as (60) shows. The use of this construction to mean ‘get scared by’ seems to be a more recent development, and is first attested in 1875. This example is presented in (61). Before that, the combination of *schrikken* with *voor* could already be used to express this meaning, as (62) shows.

- (60) *Vaste voet **schrikt** nooyt **van** 't pad.* (camphuyzen, Sticht. Rymen 222, 1624)
“A steady foot never slips off the path.”
- (61) *Hoe de soldaat haast hoopt dat er een oorlog komt, en hoe zij daarvan **schrikt**.* (quack, Stud. 267, 1875)
“How the soldier almost hopes there will be war, and how it **scares** them.”

- (62) *Goden ... voor dewelcke uwe vaders niet geschrickt en hebben.* (Statenb., Deut. 32, 17, ed. 1688)
 “Gods ... **of which** your fathers would not be **scared**.”

5.4.12 Voorzien van ‘provide with’

The earliest sentences with *voorzien* show two constructions: one with the person to whom something is provided marked with the dative and that which is provided as a direct object, and one with the provided items in a prepositional phrase with *van*. Examples of both are given in (63) and (64), respectively.

- (63) *dies machmen merken wel in dien dat die sceppre heuet vorsien elker creaturen thare alst herlic ende nuttelic ware.* (Nat.Bl.D p. 118, r. 40-43, 1287)
 “Hieruit kan men wel begrijpen aldus dat de schepper ieder schepsel heeft voorzien van het hare, als het betamelijk en nuttig zou zijn.” [From this it may well be understood that the Creator has provided every creature with its own, if it were proper and useful.]
- (64) *Jnde dar+umbe+dat got den minsche also sere mint. dat hine vorsiin wille van al des hi bedarf. so heuet hi den minsche gegeuen ene manire van sterkheit van sile di heit gedinchenis.* (Moraalb. p. 355, r. 10-12, 1270-1290)
 “En omdat God de mens zo zeer bemint dat hij hem **voorzien** wil **van alles wat hij nodig heeft**, heeft hij de mens innerlijke kracht gegeven, die heet *gedinchenis* (‘geheugen’).” [And because God loves man so much that he wants to **provide** him **with everything he needs**, he has given man inner strength, which is called *gedinchenis* (“memory”).]

Combinations with *in* and *tot* are also found, but there is a clear difference in meaning with these sentences (i.e., ‘arrange/take care of’); the former is still attested in modern Dutch. There was, however, a period in which *met* was used to express the same meaning as *van*. An example can be found in (65). This construction is found from the early 16th century to the end of the 19th century, alongside *voorzien van*.

- (65) *De kosten der mede, die toch ten besten niet beslaeght is, verdienen niet gerekend te worden. Wy zullen op 't aenstaende voorjaer trachten U Ed. Gestr. met wat beters te voerzien.* (hooft, Br. 3, 101, 1636)
 “De kosten daarvan, die niet gunstig zijn uitgevallen, dienen niet verrekend te worden. Wij zullen het aanstaende voorjaar trachten U Edele Gestreng **van wat beters te voorzien**.” [The costs of that, which have not turned out favorably, should not be settled. We will try to **provide** you, Sir, **with something better** next spring.]

5.5. Discussion

This chapter presented an overview of the development of the Dutch case system, prepositions and prepositional predicates. Twelve predicates were selected to investigate more

closely using dictionary data from the Old Dutch Dictionary, Early Middle Dutch Dictionary, Middle Dutch Dictionary, and Dutch Dictionary (which covers (early) modern Dutch).

There was quite a lot of variation in the developmental patterns of these predicates. For some, there is data available showing the use of dative and genitive case marking, while others are found with prepositional object from a very early period (e.g., *verhalen aan/op* in the 13th century). There are predicates that have always used the same preposition (e.g., *verzetten*), while others shifted around for a while (e.g., *beschermen*). An overview of the main findings is presented in Table 5.2. Each row describes one of the predicates included in this diachronic study. It mentions whether it was included because of its high or low intra preposition group dissimilarity score, the case it originally governed (though that information could not be found for the majority of the predicates), and how the object was marked throughout history. The first century in which a form or a combination of different forms could be found in the data is also noted.

Table 5.2

Summary of the development of the twelve predicates discussed in the current chapter.

| Predicate | High/low intra P group dissimilarity score | Original case | Object marking (century first attested in data) |
|---|--|------------------|---|
| <i>Azen op</i> ‘have one’s eye on’ | Low (0.85) | <i>Not found</i> | naar (13 th) → op (18 th) |
| <i>Lijken op</i> ‘resemble’ | High (1.80) | dat | dat/tegen (15 th) → Ø (17 th) → naar (18 th) → op (19 th) |
| <i>Verhalen op</i> ‘recover (smth.) from’ | High (1.80) | <i>Not found</i> | op/aan (13 th) → op/aan/tegen/in/van/met (<i>unclear</i>) → op (16 th) |
| <i>Vertrouwen op</i> ‘trust/rely on’ | Low (0.88) | dat | [dat/op] _{betrouwen/getrouwen} (13 th) → [op/in] _{vertrouwen} (16 th) |
| <i>Beschermen tegen</i> ‘protect against’ | Low (0.40) | gen | gen → weder/van (10 th) → van/tegen (13 th) → van/tegen/voor (14 th) → tegen/voor (16 th) |
| <i>Kunnen tegen</i> ‘can stand’ | High (1.70) | <i>Not found</i> | tegen (19 th) |
| <i>Verweren tegen</i> ‘defend against’ | Low (0.40) | gen | gen → gen/van/tegen/voor (13 th) → tegen (14 th) → tegen/van (17 th) → tegen (19 th) |
| <i>Verzetten tegen</i> ‘resist’ | High (1.70) | <i>Not found</i> | tegen (13 th) |
| <i>Dromen van</i> ‘dream of’ | Low (1.20) | <i>Not found</i> | van (17 th) |
| <i>Overtuigen van</i> ‘convince of’ | Low (1.20) | gen | gen → Ø (16 th) → van (17 th) |
| <i>Schrikken van</i> ‘get scared by’ | High (1.80) | <i>Not found</i> | voor (17 th) → van (19 th) |
| <i>Voorzien van</i> ‘provide with’ | High (1.75) | <i>Not found</i> | Ø/van (13 th) → van/met (16 th) → van (19 th) |

Van der Horst (2008) offers data that shows that the group of prepositional predicates only really started developing in late Middle Dutch (15th century), and became a group of a significant size from the 16th century onwards. It is therefore very surprising that the data for the twelve predicates discussed here shows a use of prepositions quite early on. Only for five of these predicates a case marking construction is listed, while the other predicates are found in prepositional constructions from their earliest attestations. One might then assume that the data for this last group is not as old as the former group, but that is also not the case. *Kunnen tegen* and *schrikken van* aside, which have both very recently (19th century) developed the here intended meaning, all predicates have data from the 13th century. The case marked constructions, on the other hand, are found until the 17th century in written language, although they had disappeared from spoken language. It seems, therefore, that there is a lot of variation in when these predicates developed into prepositional predicates, and when they lost case marking. Based on literature, it seems reasonable to assume that, even though it is not present in this data, at some point the predicates did govern a case marked object, rather than a prepositional phrase. The conclusion can be drawn, therefore, that some of the data here is simply not old enough.

With the exception of *lijken + dative*, all case marking constructions are found with predicates with a low intra preposition group dissimilarity score (i.e., those that were most similar to others in the same group). This might point to a pattern, and to the idea that verbs that governed the same case are more similar to each other, but there is too little data so make such claims.

For the predicates on which case marking data is available, it is directly in line with expectations based on literature. Van der Horst (2008) and Vandeweghe (2011) discussed how, for example, prepositional phrases with *van* developed from genitive phrases, phrases with *naar* from accusative, and phrases with *met* from dative. In the same vein, this dictionary data shows that predicates that now govern the same prepositions, also governed the same case. Predicates with *op* used to govern a dative object, and predicates that now govern a prepositional phrase with *tegen* and *van*, both used to govern a genitive object. Vandeweghe hypothesises that the choice for a certain preposition is (partially) semantically motivated. Prepositions like *van* and *uit*, for example have more “affinity” with the genitive meaning of ‘origin’, and are thus used to replace genitive constructions. That is also the case here with the group of predicates that govern a *van* phrase. As to whether this observation holds on a larger scale, more data is needed.

To conclude, the data shows much variation in when and how predicates developed over time. There seems to be a trend that case marking is found with predicates that are more similar to others from the same preposition group, but more data is needed for a definite conclusion on that topic. Secondly, the data shows that predicates that now govern the same preposition developed from constructions that governed the same case. In the same way the cross-linguistic data showed lower dissimilarity within groups, this data also shows that predicates that govern the same preposition are more similar to each other. This similarity also seems to be semantically motivated. Further investigation would be interesting in order to see whether this is also true for other preposition groups.

Chapter 6. Conclusion

In this thesis, I set out to investigate semantic groups among Dutch prepositional predicates. Chapters 3 to 5 described the three studies that were carried out to investigate the semantics of the preposition in a prepositional predicates. Much literature on them assumes the preposition is meaningless, and the choice of preposition is arbitrary.

The semantic similarity analysis using Latent Semantic Analysis, described in chapter 3, yielded no information for the research question. It is likely that the method would benefit from more data and more data preparation. Patterns such as those found by Höllein (2019) for German might be hiding in the collected data, but finding them requires a more refined analysis. Suggestions were made in section 3.4 to further prepare the data for a later study. Investigation only content words would be one option, but another way to investigate this would be to annotate semantic roles based on corpus data (thus providing quantitative data to support the groups found by Loonen (2003).

In Chapter 4, a cross-linguistic study was conducted to compare Dutch predicates with English and German ones, providing two main points. First, predicates group together similarly across language, i.e. when Dutch predicates use the same preposition, the English and German equivalents tend to do the same. This provides clear evidence to the fact that preposition selection is *not* random. Second, o this study found that the English and German prepositions that were used for the translations of Dutch verbs, occupied the same semantic space as the Dutch preposition. For example, Dutch verbs that use *tegen* were translated with English verbs using *against* and German verbs using *gegen*. This goes against the statements in Hoeksema (2014), Ackema (2015), and Saint-Dizier (2007), who describe how different languages use different prepositions.

Finally, the diachronic study in Chapter 5 shows that there is much variation in the way that the investigated predicates evolved over time. The study would benefit from a larger set of predicates and corpus data in addition to the dictionary data that was discussed. Nonetheless, some patterns appear to emerge: prepositions that governed the same case in history, now use the same preposition. Which preposition that is, does also correspond to what is described in literature: verbs with genitive objects started using the preposition *van*, for example. Interestingly, two verbs that used genitive object marking and then used *van*, later changed their preposition to *tegen* (viz. *beschermen* ‘protect’ and (*zich*) *verweren* ‘defend (oneself)’). This could be due to assimilation with the way semantically related verbs expressed their object (*verzetten* ‘resist’, for example, used *van* from very early on—the 13th century, at least), but confirming that theory requires further investigation.

It can be concluded that claims by authors like Heringer (1968, p. 434-435) that “no semantic significance” can be ascribed to the preposition are proven untrue, both by other literature and the data presented in this thesis. The idiosyncratic nature of preposition choice is also disproven. While some differences exist, there is clear overlap in the clustering of prepositions across language boundaries. Some meaning is has been lost during the grammaticalisation of the construction, but not all meaning, and not for all predicates. Therefore, while most prepositions in prepositional predicates don’t carry their original spatial or temporal meaning, they *are* selected based on the semantics of both the verb and (a metaphorical interpretation of) that preposition.

References

- Ackema, P. (2015). 9. *Arguments and Adjuncts*. De Gruyter EBooks, 246–274.
<https://doi.org/10.1515/9783110377408.246>
- Anderson, L. B. (1982). The “perfect” as a universal and as a language-particular category. In P. J. Hopper (Ed.), *Tense-Aspect: Between Semantics & Pragmatics (Typological Studies in Language)* (pp. 227–264). Amsterdam, The Netherlands: John Benjamins.
<https://doi.org/10.1075/tsl.1.16and>
- Beliën, M. (2021, januari). 9.1 Voorzetsels (preposities) (versie 3.0). *Algemene Nederlandse Spraakkunst*. <https://e-ans.ivdnt.org/topics/pid/topic-16249634411997549>. (geraadpleegd 27 mei 2023).
- Beliën, M. (2021, januari). 9.4.2 Een overzicht van voorzetseluitdrukkingen (versie 3.0). *Algemene Nederlandse Spraakkunst*. <https://e-ans.ivdnt.org/topics/pid/topic-16249639081368946>.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. Sebastopol (CA), USA: O’Reilly Media, Inc.
- Borchmann, L. (2017). *Stress1* branch of *skikit-learn*, GitHub, <https://github.com/Borchmann/scikit-learn/tree/stress1>
- Borg, I., & Groenen, P. (2005). Modern Multidimensional Scaling. *Springer eBooks*.
<https://doi.org/10.1007/0-387-28981-x>
- Botwinik-Rotem, I. (2004). *The Category of P. Features, Projections, Interpretation*. Ph.D. Thesis, Tel Aviv University.
- Botwinik-Rotem, I. (2011). *PP-verbs: The phenomenon of obligatory Ps*. Ms., Tel Aviv University.
- Bouma, G., van Noord, G. J. M. and Malouf, R. 2001. Alpino: Wide-coverage computational analysis of Dutch. in Daelemans, W., Simaan, K., Veenstra, J. and Zavrel, J. (eds.). *Computational Linguistics in the Netherlands 2000*. Amsterdam: Rodopi, p. 45-59 15 p. (LANGUAGE AND COMPUTERS : STUDIES IN PRACTICAL LINGUISTICS)
- Bradford, R. B. (2008). An empirical study of required dimensionality for large-scale latent semantic indexing applications. *CIKM ’08: Proceeding of the 17th ACM Conference on Information and Knowledge Mining - CIKM ’08*, 153–162.
<https://doi.org/10.1145/1458082.1458105>
- Broekhuis, H. (2004). Het voorzetselvoorwerp. *Nederlandse Taalkunde*, 9(2), 97–131.
- Broekhuis, H. (2007). Worstelen met het voorzetselvoorwerp: opmerkingen bij Ina Schermer-Vermeers. *Nederlandse Taalkunde*, 12(4), 351–358.
- Broekhuis, H. (2014). Dubbel-voorzetselobjectconstructies? In F. Van De Velde, H. Smessaert, F. Van Eynde, & S. Verbrugge (Eds.), *Patroon en Argument: Een dubbelfestbundel bij het emeritaat van William Van Belle en Joop van der Horst* (pp. 103–114). Leuven: Universitaire Pers Leuven.
- Broekhuis, H., Corver, N., & Vos, R. (2015). *Syntax of Dutch: Verbs and Verb Phrases*. Amsterdam: Amsterdam University Press.

- De Swart, P. (2014). Prepositional inanimates in Dutch: A paradigmatic case of Differential Object Marking. *Linguistics*, 52(2). <https://doi.org/10.1515/ling-2013-0069>
- Deerwester, S., Dumais, S., Furnas, G., Landauer, T., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407. Retrieved from [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6%3C391::AID-ASII%3E3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6%3C391::AID-ASII%3E3.0.CO;2-9)
- Duden online. (n.d.). Retrieved June 27, 2020, from <https://www.duden.de/>
- Duinhoven, A. (1989). Het voorzetselvoorwerp: Een zinspatroon in wording. *De Nieuwe Taalgids*, 82, 40–54.
- Engel, U., & Schumacher, H. (1976). *Kleines Valenzlexikon deutscher Verben*. Tübingen: Narr.
- Engel, U., Schumacher, H., & Riebschläger, H. (1978). Kleines Valenzlexikon deutscher Verben. *Info DaF. Informationen Deutsch Als Fremdsprache*, 5(3), 19–20. <https://doi.org/10.1515/infodaf-1978-050311>
- Faber, I., Rem, M., & De Hoop, H. (2022). From visual perception to inferential evidentiality: The case of Dutch eruitzien ‘look.’ *Nederlandse Taalkunde*, 27(2), 229–265. <https://doi.org/10.5117/nedtaa2022.2.007.hoop>
- Firth, E. R. (1957). *Studies in Linguistic Analysis*. Oxford: Blackwell.
- Gefen, D., Endicott, J. N., Fresneda, J. E., Miller, J. A., & Larsen, K. (2017). A Guide to Text Analysis with Latent Semantic Analysis in R with Annotated Code Studying Online Reviews and the Stack Exchange Community. *Communications of the Association for Information Systems*, 41, 450–496. <https://doi.org/10.17705/1cais.04121>
- German/Grammar/Prepositions and Postpositions. (2020, February 13). Retrieved August 6, 2020, from https://en.wikibooks.org/wiki/German/Grammar/Prepositions_and_Postpositions
- Haeringen, C. B. (1956). *Nederlands tussen Duits en Engels*. Den Haag: Servire. Retrieved from https://www.dbnl.org/tekst/haer001nede01_01/index.php
- Haeseryn, W., Romijn, K., Geerts, G., De Rooij, J., & Van Den Toorn, M. (1997). *Algemene Nederlandse Spraakkunst* (2nd ed.). Groningen/Deurne, Netherlands: Martinus Nijhoff uitgevers/Wolters Plantyn.
- Harris, Z. S. (1954). Distributional Structure. *WORD*, 10(2–3), 146–162. <https://doi.org/10.1080/00437956.1954.11659520>
- Heringer, H. J. (1968). Präpositionale Ergänzungsbestimmungen im Deutschen. *Zeitschrift Für Deutsche Philologie*, 87, 426–457.
- Historische woordenboeken. (n.d.). Retrieved from <https://gtb.ivdnt.org/search/>
- Hoeksema, J. (2014). De plaats van het voorzetselvoorwerp. *Nederlandse Taalkunde*, 19(2), 221–244. <https://doi.org/10.5117/nedtaa2014.2.hoek>
- Hogehout-Mulder, M. (2007). *Cursus Middelnederlands*. Groningen: Wolters-Noordhoff.
- Höllein, D. (2019). *Präpositionalobjekt vs. Adverbial: Die semantischen Rollen der Präpositionalobjekte*. Walter de Gruyter GmbH & Co KG.
- Hopper, P. J., & Traugott, E. C. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods*, 42(3), 643–650. <https://doi.org/10.3758/BRM.42.3.643>
- Kintsch, W., McNamara, D., Dennis, S., & Landauer, T. (2007). LSA and meaning: In theory and application. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 467–180). Mahwah (NJ)/London, USA/UK: Lawrence Erlbaum Associates, Publishers.
- Koeman, J. H., & Rea, W. J. (2014). How Does Latent Semantic Analysis Work? A Visualisation Approach. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1402.0543>
- Kruskal, J. B., & Wish, M. (1978). Multidimensional Scaling. *SAGE Publications, Inc. eBooks*. <https://doi.org/10.4135/9781412985130>
- Kruskal, J.B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29: 115–129.
- Kwantes, P. (2005). Using context to build semantics. *Psychonomic Bulletin & Review*, 12(4), 703–710. <https://doi.org/10.3758/BF03196761>
- Lakoff, G. (1987). *Women, Fire and Dangerous Things. What Categories Reveal about the Mind*. Chicago, United States of America: University of Chicago Press.
- Landauer, T. (2007). LSA as a theory of meaning. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 3–34). Mahwah (NJ)/London, USA/UK: Lawrence Erlbaum Associates, Publishers.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato’s problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211–240. Retrieved from <http://www.stat.cmu.edu/~cshalizi/350/2008/readings/Landauer-Dumais.pdf>
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259–284. Retrieved from <http://lsa.colorado.edu/papers/dp1.LSAintro.pdf>
- Landauer, T., & Dumais, S. (2008). Latent semantic analysis. *Scholarpedia*, 3(11), 4356. <https://doi.org/10.4249/scholarpedia.4356>
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Volume I: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Lassy Groot-corpus (Version 4.0) (2016) [WR-P-P-G]. Available at the Dutch Language Institute: <http://hdl.handle.net/10032/tm-a2-f6>
- Loonen, L. J. M. (2003). *Stante pede gaande van dichtbij langs AF bestemming @* (dissertation). Utrecht University.
- Ministerie van Openbaar Onderwijs. (1949). *Codex van de spraakkundige terminologie*.
- Odiijk, J., Van Noord, G., Kleiweg, P., & Tjong Kim Sang, E. (2017). The Parse and Query (PaQu) Application. In J. Odiijk & A. Van Hessen (Eds.), *CLARIN in the Low Countries* (pp. 281–297). London, England: Ubiquity Press. <https://doi.org/10.5334/bbi.23>
- Oxford University Press. (n.d.). OED Online. Retrieved June 22, 2020, from <https://www.oed.com/>

- Paardekooper, P. (1986). *Beknopte ABN-syntaxis*. Eindhoven, Netherlands: Uitgave in eigen beheer.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). [Scikit-learn: Machine Learning in Python](#), *JMLR 12*, pp. 2825-2830.
- Porter, M. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130–137. <https://doi.org/10.1108/00330330610681286>
- Quesada, J. (2007). Creating you own LSA spaces. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 71–88). Mahwah (NJ)/London, USA/UK: Lawrence Erlbaum Associates, Publishers.
- Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50).
- Řehůřek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2).
- Saint-Dizier, P. (2006). *Syntax and Semantics of Prepositions*. Springer Science & Business Media.
- Schermer-Vermeer, E.C. (1990). Het voorzetselvoorwerp en het begrip “vorm.” *De Nieuwe Taalgids*, 83, 238–247.
- Schermer-Vermeer, I. (2005). Worstelen met het voorzetselvoorwerp. *Nederlandse Taalkunde*, 11(2), 146–167.
- Schermer-Vermeer, I. (2007). Worstelen rond het voorzetselvoorwerp: een reactie op de opmerkingen van Hans Broekhuis. *Nederlandse Taalkunde*, 12(4), 358–365.
- Schofield, A., & Mimno, D. (2016). Comparing Apples to Apple: The Effects of Stemmers on Topic Models. *Transactions of the Association for Computational Linguistics*, 4, 287–300. https://doi.org/10.1162/tacl_a_00099
- Somers, H. L. (1984). On the validity of the complement-adjunct distinction in valency grammar. *Linguistics*, 22(4). <https://doi.org/10.1515/ling.1984.22.4.507>
- Stoett, F.A. (1977). *Middelnederlandsche spraakkunst. Syntaxis*. Den Haag: Martinus Nijhoff.
- Van Bart, P., & Sturm, A. (1987). *Zinsanalyse*. Leiden: Martinus Nijhoff.
- Van Dale. (n.d.). Retrieved from <https://www.vandale.nl/opzoeken>
- Van Den Toorn, M. (1971). Het voorzetselvoorwerp als nominale constituent. *Studia Neerlandica*, 6, 114–130.
- Van Der Horst, J. (2008). *Geschiedenis van de Nederlandse Syntaxis*. Leuven: Universitaire Pers Leuven.
- Van Der Sijs, N. (2010). Etymologiebank. Retrieved August 4, 2020, from <http://etymologiebank.ivdnt.org/>
- Van Driel, H., Van Den Toorn, M., & Vullings, H. (1978). Intuities omtrent het voorzetselvoorwerp. *Gramma*, 2, 37–50.
- Van Haeringen, C. (1956). *Nederlands tussen Duits en Engels*. Den Haag, Netherlands: Servire.

- Van Heeswijck, A. (1966). Wat is het nu eigenlijk: voorwerp of bepaling? *Nova Vetera*, 43(3), 161–168.
- Van Noord, G., Schuurman, I., & Vandeghinste, V. (2006). Syntactic Annotation of Large Corpora in STEVIN. (N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk, & D. Tapias, Eds.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*.
- Van Noord, G. (2009). Huge parsed corpora in Lassy. In Frank van Eynde, Anette Frank, Koenraad De Smedt, and Gertjan van Noord, editors, *Proceedings of the Seventh International Workshop on Treebanks and Linguistic Theories (TLT 7)*, number 12 in LOT Occasional Series, pages 115–126, Utrecht, The Netherlands, 2009. Netherlands Graduate School of Linguistics.
- Vandeweghe, W. (2015). Het voorzetselvoorwerp als boomende categorie. In S. Lestrade, P. De Swart, & L. Hogeweg (Eds.), *Addende: Artikelen voor Ad Foolen* (pp. 453–470). Nijmegen: Radboud Universiteit Nijmegen.
- Vandeweghe, W. (2020, March 4). re: Linguïstisch miniatuurtje CLXVIII: Wat staat er op een ontmoeting met Ad Foolen? [Comment]. Retrieved from <https://www.neerlandistiek.nl/2020/03/linguïstisch-miniatuurtje-clxviii-wat-staat-er-op-een-ontmoeting-met-ad-foolen/#comment-21890>
- Wälchli, B. (2010). Similarity Semantics and Building Probabilistic Semantic Maps from Parallel Texts. *Linguistic Discovery*, 8(1). <https://doi.org/10.1349/ps1.1537-0852.a.356>
- Wälchli, B., & Cysouw, M. (2012). Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics*, 50(3). <https://doi.org/10.1515/ling-2012-0021>
- Wikiwörterbuch: ein Wiki-basiertes freies Wörterbuch. (n.d.). Retrieved July 28, 2020, from <https://de.wiktionary.org/wiki/Wiktionary:Hauptseite>
- Zato, Z. (2014). Qualia Structure in Spanish Prepositional Verbs: When the verb resorts to a preposition. *Borealis: An International Journal of Hispanic Linguistics*. <https://doi.org/10.7557/1.3.1.2752>
- Zwaan, F. (1972). Het voorzetselvoorwerp. *Levende Talen*, 347–350.
- Zwarts, J. (2005). *The case of prepositions: Government and compositionality in German PPs*.
- Zwarts, J. (2014). Voorwerpen met voorzetsels. *Nederlandse Taalkunde*, 19(2), 245–247. <https://doi.org/10.5117/nedtaa2014.2.zwar>

Appendices

Appendix A: List of selected predicates and frequencies

| # | Preposition group | Predicate | Absolute frequency | Relative frequency (per 10,000 sentences) |
|----|-------------------|----------------------|--------------------|---|
| 1 | | danken aan | 6,094 | 4.07 |
| 2 | | (zich) hechten aan | 2,751 | 1.84 |
| 3 | | liggen aan | 7,904 | 5.28 |
| 4 | | lijden aan | 2,941 | 1.96 |
| 5 | aan | zich onderwerpen aan | 2,035 | 1.36 |
| 6 | | zich storen aan | 990 | 0.66 |
| 7 | | verbinden aan | 2,303 | 1.54 |
| 8 | | voldoen aan | 7,536 | 5.03 |
| 9 | | zich wagen aan | 2,197 | 1.47 |
| 10 | | wijten aan | 3,772 | 2.52 |
| 11 | | | berusten in | 203 |
| 12 | | betrekken in | 304 | 0.20 |
| 13 | | delen in | 1,991 | 1.33 |
| 14 | | geloven in | 8,014 | 5.35 |
| 15 | in | zich schikken in | 157 | 0.10 |
| 16 | | slagen in | 12,946 | 8.65 |
| 17 | | zich verdiepen in | 1,057 | 0.71 |
| 18 | | zich verheugen in | 204 | 0.14 |
| 19 | | volharden in | 240 | 0.16 |
| 20 | | voorzien in | 1,552 | 1.04 |
| 21 | | | gissen naar | 167 |
| 22 | naar | luisteren naar | 6,568 | 4.39 |
| 23 | | refereren naar | 154 | 0.10 |
| 24 | | smachten naar | 135 | 0.09 |
| 25 | | smaken naar | 621 | 0.41 |
| 26 | | snakken naar | 312 | 0.21 |
| 27 | | streven naar | 2,550 | 1.70 |
| 28 | | verlangen naar | 886 | 0.59 |
| 29 | | vragen naar | 2,277 | 1.52 |
| 30 | | zoeken naar | 8,368 | 5.59 |
| 31 | | zich bekommeren om | 853 | 0.57 |
| 32 | | geven om | 377 | 0.25 |
| 33 | | lachen om | 1,261 | 0.84 |
| 34 | om | malen om | 373 | 0.25 |
| 35 | | rouwen om | 248 | 0.17 |
| 36 | | smeken om | 411 | 0.27 |
| 37 | | treuren om | 132 | 0.09 |

| | | | | |
|----|-------|----------------------|--------|-------|
| 38 | | verzoeken om | 160 | 0.11 |
| 39 | | vragen om | 7,352 | 4.91 |
| 40 | | antwoorden op | 1,252 | 0.84 |
| 41 | | azen op | 971 | 0.65 |
| 42 | | zich/iets baseren op | 5,718 | 3.82 |
| 43 | | zich concentreren op | 2,156 | 1.44 |
| 44 | op | doelen op | 1,246 | 0.83 |
| 45 | | lijken op | 7,237 | 4.83 |
| 46 | | zich beroepen op | 1,020 | 0.68 |
| 47 | | verhalen op | 551 | 0.37 |
| 48 | | vertrouwen op | 810 | 0.54 |
| 49 | | wijzen op | 10,578 | 7.06 |
| 50 | | zich beraden over | 320 | 0.21 |
| 51 | | beslissen over | 3,302 | 2.21 |
| 52 | | discussiëren over | 1,225 | 0.82 |
| 53 | | klagen over | 4,123 | 2.75 |
| 54 | over | onderhandelen over | 2,109 | 1.41 |
| 55 | | zich ontfermen over | 307 | 0.21 |
| 56 | | oordelen over | 1,034 | 0.69 |
| 57 | | piekeren over | 300 | 0.20 |
| 58 | | waken over | 574 | 0.38 |
| 59 | | zwijgen over | 1,090 | 0.73 |
| 60 | | beschermen tegen | 1,703 | 1.14 |
| 61 | | zich keren tegen | 1,404 | 0.94 |
| 62 | | kunnen tegen | 2,047 | 1.37 |
| 63 | | pleiten tegen | 155 | 0.10 |
| 64 | tegen | protesteren tegen | 1,062 | 0.71 |
| 65 | | strijden tegen | 769 | 0.51 |
| 66 | | vechten tegen | 1,518 | 1.01 |
| 67 | | zich verweren tegen | 191 | 0.13 |
| 68 | | zich verzetten tegen | 3,107 | 2.07 |
| 69 | | zich wapenen tegen | 332 | 0.22 |
| 70 | | behoren tot | 12,398 | 8.28 |
| 71 | | zich beperken tot | 2,110 | 1.41 |
| 72 | | besluiten tot | 883 | 0.59 |
| 73 | | bijdragen tot | 1,934 | 1.29 |
| 74 | tot | dwingen tot | 1,660 | 1.11 |
| 75 | | herleiden tot | 1,290 | 0.86 |
| 76 | | leiden tot | 20,209 | 13.50 |
| 77 | | spreken tot | 857 | 0.57 |
| 78 | | zich verlagen tot | 505 | 0.34 |
| 79 | | verleiden tot | 706 | 0.47 |
| 80 | | betichten van | 819 | 0.55 |
| 81 | van | dromen van | 4,700 | 3.14 |
| 82 | | genieten van | 12,639 | 8.44 |
| 83 | | getuigen van | 2,437 | 1.63 |

| | | | | |
|----|------|-----------------------|--------|-------|
| 84 | | (zich) ontdoen van | 624 | 0.42 |
| 85 | | overtuigen van | 930 | 0.62 |
| 86 | | schrikken van | 1,515 | 1.01 |
| 87 | | verschillen van | 2,145 | 1.43 |
| 88 | | (zich) verzekeren van | 996 | 0.67 |
| 89 | | voorzien van | 4,278 | 2.86 |
| 90 | | bedanken voor | 615 | 0.41 |
| 91 | | behoeden voor | 879 | 0.59 |
| 92 | | ijveren voor | 1,727 | 1.15 |
| 93 | | kiezen voor | 23,556 | 15.73 |
| 94 | voor | zich lenen voor | 514 | 0.34 |
| 95 | | pleiten voor | 7,946 | 5.31 |
| 96 | | voelen voor | 1,617 | 1.08 |
| 97 | | vrezen voor | 3,726 | 2.49 |
| 98 | | waarschuwen voor | 728 | 0.49 |
| 99 | | zorgen voor | 44,948 | 30.02 |

Appendix B: Frequencies of predicates for LSA

The table below shows, for each of the predicates, the number of sentences found in the corpus, the number of sentences with multiple instances of the preposition in question (i.e., the number of sentences removed from the dataset), and the number of sentences that the analysis was carried out with.

| # | Predicate | Total number of hits in corpus | More than one preposition (%) | Number of sentences remaining |
|----|----------------------|--------------------------------|-------------------------------|-------------------------------|
| 1 | Danken aan | 5,898 | 533 (9.04%) | 5,365 |
| 2 | (Zich) hechten aan | 2,714 | 253 (9.32%) | 2,461 |
| 3 | Liggen aan | 7,696 | 775 (10.07%) | 6,921 |
| 4 | Lijden aan | 2,913 | 268 (9.2%) | 2,645 |
| 5 | Zich onderwerpen aan | 2,027 | 134 (6.61%) | 1,893 |
| 6 | Zich storen aan | 977 | 98 (10.03%) | 879 |
| 7 | Verbinden aan | 2,293 | 143 (6.24%) | 2,150 |
| 8 | Voldoen aan | 3,838 | 286 (7.45%) | 3,552 |
| 9 | Zich wagen aan | 2,189 | 110 (5.03%) | 2,079 |
| 10 | Wijten aan | 3,625 | 507 (13.99%) | 3,118 |
| 11 | Berusten in | 202 | 48 (23.76%) | 154 |
| 12 | Betrekken in | 304 | 55 (18.09%) | 249 |
| 13 | Delen in | 1,986 | 362 (18.23%) | 1,624 |
| 14 | Geloven in | 7,884 | 1,346 (17.07%) | 6,538 |
| 15 | Zich schikken in | 157 | 25 (15.92%) | 132 |
| 16 | Slagen in | 12,933 | 3,827 (29.59%) | 9,106 |
| 17 | Zich verdiepen in | 1,050 | 262 (24.95%) | 788 |
| 18 | Zich verheugen in | 200 | 60 (30%) | 140 |
| 19 | Volharden in | 240 | 43 (17.92%) | 197 |
| 20 | Voorzien in | 1,539 | 438 (28.46%) | 1,101 |
| 21 | Gissen naar | 8,290 | 361 (4.35%) | 7,929 |
| 22 | Luisteren naar | 166 | 8 (4.82%) | 158 |
| 23 | Refereren naar | 6,461 | 450 (6.96%) | 6,011 |
| 24 | Smachten naar | 152 | 10 (6.58%) | 142 |
| 25 | Smaken naar | 133 | 11 (8.27%) | 122 |
| 26 | Snakken naar | 611 | 33 (5.4%) | 578 |
| 27 | Streven naar | 310 | 8 (2.58%) | 302 |
| 28 | Verlangen naar | 2,517 | 142 (5.64%) | 2,375 |
| 29 | Vragen naar | 876 | 62 (7.08%) | 814 |
| 30 | Zoeken naar | 2,244 | 144 (6.42%) | 2,100 |
| 31 | Zich bekommeren om | 840 | 78 (9.29%) | 762 |
| 32 | Geven om | 372 | 61 (16.4%) | 311 |
| 33 | Lachen om | 1,258 | 48 (3.82%) | 1,210 |
| 34 | Malen om | 372 | 30 (8.06%) | 342 |

| | | | | |
|----|----------------------|--------|----------------|--------|
| 35 | Rouwen om | 244 | 18 (7.38%) | 226 |
| 36 | Smeken om | 409 | 31 (7.58%) | 378 |
| 37 | Treuren om | 129 | 11 (8.53%) | 118 |
| 38 | Verzoeken om | 160 | 9 (5.63%) | 151 |
| 39 | Vragen om | 7,288 | 508 (6.97%) | 6,780 |
| 40 | Antwoorden op | 1,250 | 142 (11.36%) | 1,108 |
| 41 | Azen op | 966 | 102 (10.56%) | 864 |
| 42 | Zich/iets baseren op | 5,580 | 687 (12.31%) | 4,893 |
| 43 | Zich concentreren op | 1,009 | 135 (13.38%) | 874 |
| 44 | Doelen op | 2,123 | 245 (11.54%) | 1,878 |
| 45 | Lijken op | 1,200 | 214 (17.83%) | 986 |
| 46 | Zich beroepen op | 7,182 | 1,065 (14.83%) | 6,117 |
| 47 | Verhalen op | 547 | 57 (10.42%) | 490 |
| 48 | Vertrouwen op | 794 | 129 (16.25%) | 665 |
| 49 | Wijzen op | 10,476 | 1,481 (14.14%) | 8,995 |
| 50 | Zich beraden over | 318 | 12 (3.77%) | 306 |
| 51 | Beslissen over | 3,286 | 110 (3.35%) | 3,176 |
| 52 | Discussiëren over | 1,202 | 95 (7.9%) | 1,107 |
| 53 | Klagen over | 4,086 | 170 (4.16%) | 3,916 |
| 54 | Onderhandelen over | 2,092 | 97 (4.64%) | 1,995 |
| 55 | Zich ontfermen over | 306 | 9 (2.94%) | 297 |
| 56 | Oordelen over | 1,027 | 54 (5.26%) | 973 |
| 57 | Piekeren over | 300 | 12 (4%) | 288 |
| 58 | Waken over | 571 | 21 (3.68%) | 550 |
| 59 | Zwijgen over | 1,084 | 74 (6.83%) | 1,010 |
| 60 | Beschermen tegen | 1,690 | 53 (3.14%) | 1,637 |
| 61 | Zich keren tegen | 1,379 | 91 (6.6%) | 1,288 |
| 62 | Kunnen tegen | 2,035 | 56 (2.75%) | 1,979 |
| 63 | Pleiten tegen | 151 | 14 (9.27%) | 137 |
| 64 | Protesteren tegen | 1,047 | 43 (4.11%) | 1,004 |
| 65 | Strijden tegen | 753 | 45 (5.98%) | 708 |
| 66 | Vechten tegen | 1,488 | 87 (5.85%) | 1,401 |
| 67 | Zich verweren tegen | 190 | 7 (3.68%) | 183 |
| 68 | Zich verzetten tegen | 3,083 | 128 (4.15%) | 2,955 |
| 69 | Zich wapenen tegen | 331 | 2 (0.6%) | 329 |
| 70 | Behoren tot | 12,344 | 499 (4.04%) | 11,845 |
| 71 | Zich beperken tot | 2,105 | 110 (5.23%) | 1,995 |
| 72 | Besluiten tot | 882 | 36 (4.08%) | 846 |
| 73 | Bijdragen tot | 1,921 | 78 (4.06%) | 1,843 |
| 74 | Dwingen tot | 1,648 | 80 (4.85%) | 1,568 |
| 75 | Herleiden tot | 1,286 | 46 (3.58%) | 1,240 |
| 76 | Leiden tot | 19,811 | 1,184 (5.98%) | 18,627 |
| 77 | Spreken tot | 849 | 42 (4.95%) | 807 |
| 78 | Zich verlagen tot | 504 | 26 (5.16%) | 478 |
| 79 | Verleiden tot | 706 | 36 (5.1%) | 670 |
| 80 | Betichten van | 809 | 342 (42.27%) | 467 |

| | | | | |
|----|-----------------------|---------|----------------|---------|
| 81 | Dromen van | 4,657 | 1,044 (22.42%) | 3,613 |
| 82 | Genieten van | 12,449 | 3,304 (26.54%) | 9,145 |
| 83 | Getuigen van | 2,395 | 888 (37.08%) | 1,507 |
| 84 | (Zich) ontdoen van | 617 | 227 (36.79%) | 390 |
| 85 | Overtuigen van | 929 | 454 (48.87%) | 475 |
| 86 | Schrikken van | 1,506 | 442 (29.35%) | 1,064 |
| 87 | Verschillen van | 2,132 | 780 (36.59%) | 1,352 |
| 88 | (Zich) verzekeren van | 991 | 419 (42.28%) | 572 |
| 89 | Voorzien van | 4,248 | 1,581 (37.22%) | 2,667 |
| 90 | Bedanken voor | 609 | 77 (12.64%) | 532 |
| 91 | Behoeden voor | 876 | 71 (8.11%) | 805 |
| 92 | Ijveren voor | 1,694 | 326 (19.24%) | 1,368 |
| 93 | Kiezen voor | 23,049 | 2,915 (12.65%) | 20,134 |
| 94 | Zich lenen voor | 512 | 46 (8.98%) | 466 |
| 95 | Pleiten voor | 7,853 | 1,375 (17.51%) | 6,478 |
| 96 | Voelen voor | 1,613 | 161 (9.98%) | 1,452 |
| 97 | Vrezen voor | 3,705 | 258 (6.96%) | 3,447 |
| 98 | Waarschuwen voor | 725 | 68 (9.38%) | 657 |
| 99 | Zorgen voor | 44,550 | 4,424 (9.93%) | 40,126 |
| - | Total | 309,018 | 38372 (12.02%) | 270,646 |

Appendix C: Keywords based on LSA

The top five keywords for all 99 predicates before stop word removal:

| | kw1 | kw2 | kw3 | kw4 | kw5 |
|-----------------|---------|---------|----------|-------------|------------|
| danken aan | van | het | dat | die | eens |
| hechten aan | van | het | eens | hecht | dat |
| liggen aan | van | het | ligt | eens | die |
| lijden aan | eens | van | lijdt | het | die |
| onderwerpen aan | eens | van | het | onderworpen | die |
| storen aan | van | het | een | die | zich |
| verbinden aan | van | eens | het | verbonden | zijne |
| voldoen aan | van | voldoen | het | eens | die |
| wagen aan | eens | van | zich | het | waagt |
| wijten aan | van | het | dat | eens | die |
| berusten in | van | het | berust | berusten | een |
| betrekken in | het | van | een | betrekken | zijn |
| delen in | van | het | eens | die | ook |
| geloven in | van | eens | gelooft | het | die |
| schikken in | van | een | zich | het | die |
| slagen in | van | eens | het | slaagde | zijn |
| verdiepen in | van | het | eens | zich | verdiepte |
| verheugen in | van | eens | zich | verheugen | het |
| volharden in | van | het | een | volhardt | volharden |
| voorzien in | eens | van | voorzien | het | voor |
| zoeken naar | van | eens | het | zoeken | die |
| gissen naar | van | het | gissen | blijft | naar |
| luisteren naar | van | eens | het | die | luisteren |
| refereren naar | van | het | eens | refereert | die |
| smachten naar | eens | van | het | die | smacht |
| smaken naar | smaakte | het | eens | van | meer |
| snakken naar | snakken | eens | van | het | die |
| streven naar | van | eens | het | streeft | streven |
| verlangen naar | van | eens | het | die | verlangen |
| vragen naar | van | eens | het | zijne | vragen |
| bekommeren om | van | het | eens | zich | die |
| geven om | het | van | eens | die | dat |
| lachen om | van | lachen | het | eens | die |
| malen om | niets | het | maar | eens | van |
| rouwen om | van | rouwt | eens | rouwen | het |
| smeken om | eens | smeekte | van | het | smeken |
| treuren om | treuren | van | het | niets | eens |
| verzoeken om | van | eens | verzocht | het | dat |
| vragen om | eene | van | het | vraagt | die |
| antwoorden op | van | eens | het | antwoorden | antwoordde |
| azen op | aast | van | eens | het | azen |

| | | | | | |
|--------------------|---------|-------------|--------------|--------------|----------------|
| baseren op | van | het | eens | die | dat |
| beroepen op | van | het | eens | zich | die |
| concentreren op | van | het | concentreren | eens | zich |
| doelen op | van | het | die | eens | doelt |
| lijken op | eens | van | het | lijkt | die |
| verhalen op | van | het | verhalen | eens | die |
| vertrouwen op | van | eens | het | dat | vertrouwen |
| wijzen op | van | eene | het | wijst | die |
| beraden over | van | beraden | het | zich | eens |
| beslissen over | van | eens | het | besliste | die |
| discussieren over | van | het | eens | discussieren | gediscussieerd |
| klagen over | van | het | klagen | eens | die |
| onderhandelen over | van | eens | het | met | onderhandelen |
| ontfermen over | van | zich | ontfermen | het | eens |
| oordelen over | van | eens | oordelen | het | die |
| piekeren over | van | het | eens | niets | piekert |
| waken over | van | het | waken | eens | dat |
| zwijgen over | van | het | eens | die | zwijgen |
| beschermen tegen | van | beschermen | het | eens | die |
| keren tegen | van | het | zich | eens | die |
| kunnen tegen | eens | van | kans | het | niets |
| pleiten tegen | van | het | eens | pleit | die |
| protesteren tegen | van | protesteren | het | eens | die |
| strijden tegen | van | het | eens | strijdt | strijden |
| vechten tegen | eens | van | het | die | vechten |
| verweren tegen | van | zich | eens | het | verweren |
| verzetten tegen | van | het | eens | zich | die |
| wapenen tegen | van | het | eens | wapenen | zich |
| behoren tot | van | het | behoort | eens | behoren |
| beperken tot | van | het | beperken | eens | zich |
| besluiten tot | van | eens | het | besloten | besloot |
| bijdragen tot | van | het | eens | die | dat |
| dwingen tot | eens | van | het | die | dwong |
| herleiden tot | eens | van | het | herleid | die |
| leiden tot | van | eens | het | dat | die |
| spreken tot | van | het | eens | spreekt | die |
| verlagen tot | van | het | eens | verlaagd | voor |
| verleiden tot | eens | van | het | verleiden | dat |
| betichten van | beticht | eens | het | hij | dat |
| dromen van | eens | droomt | het | dromen | die |

| | | | | | |
|------------------|-------------|------------|-----------|-------------|--------|
| genieten van | eens | het | genieten | die | met |
| getuigen van | eens | het | getuigt | dat | die |
| ontdoen van | eens | het | ontdoen | zijn | zich |
| overtuigen van | eens | het | zijn | overtuigen | dat |
| schrikken van | het | eens | dat | die | zijn |
| verschillen van | het | eens | verschilt | verschillen | die |
| verzekeren van | eens | verzekerde | zich | het | met |
| voorzien van | eens | voorzien | het | die | zijn |
| bedanken voor | van | eens | bedanken | het | die |
| behoeden voor | eens | van | het | behoeden | die |
| ijveren voor | van | eens | het | ijvert | die |
| kiezen voor | eens | van | het | koos | kiest |
| lenen voor | van | het | eens | leent | zich |
| pleiten voor | van | eens | pleit | het | die |
| voelen voor | van | eens | het | voelt | die |
| vrezen voor | van | eens | het | vreest | vrezen |
| waarschuwen voor | waarschuwen | van | eens | het | die |
| zorgen voor | van | eens | het | zorgde | zorgt |

The top five keywords for all 99 predicates after stop word removal:

| | kw1 | kw2 | kw3 | kw4 | kw5 |
|-----------------|-------------|-------------|----------------|------------|--------------|
| danken aan | danken | dankte | vooral | volgen | succes |
| hechten aan | hecht | hechten | belang | gehecht | waard |
| liggen aan | ligt | lag | basis | liggen | volgen |
| lijden aan | lijdt | lijden | leed | mensen | ziektes |
| onderwerpen aan | onderworpen | onderwerpen | werden | onderwierp | onderzoekers |
| storen aan | stoort | storen | stoorde | stoor | gestoord |
| verbinden aan | verbonden | verbinden | verbindt | naam | wel |
| voldoen aan | voldoen | moeten | eisen | alle | nieuws |
| wagen aan | waagt | wagen | gewaagd | waagde | eerst |
| wijten aan | wijten | wijt | volgen | vooral | deels |
| berusten in | berusten | berust | berustte | lot | nederlaag |
| betrekken in | betrekken | willen | jaar | project | mensen |
| delen in | delen | opgedeeld | deelde | deelt | klappen |
| geloven in | gelooft | geloven | geloof | geloofd | wel |
| schikken in | schikken | schikte | rol | lot | wel |
| slagen in | slaagde | geslaagde | slaagt | slagen | slaagden |
| verdiepen in | verdiepte | verdiepen | verdiep | wel | jaar |
| verheugen in | mag | verheugen | belangstelling | mogen | verheugt |
| volharden in | volharden | volhardt | boosheid | volhardde | blijven |
| voorzien in | voorzien | nieuw | eigen | moeten | behoeftes |
| zoeken naar | zoeken | zoekt | gezocht | zocht | nieuws |

| | | | | | |
|----------------|---------------|----------------|--------------|-------------|----------------|
| gissen naar | blijft | gissen | reden | oorzaak | alleen |
| luisteren naar | luisteren | luistert | geluisterd | luister | luisterde |
| refereren naar | refereert | refereren | refereerde | refererend | refereer |
| smachten naar | smacht | smachten | smachtend | waar | smachtte |
| smaken naar | smaakte | smaken | eerst | echt | wel |
| snakken naar | snakken | mensen | spelers | einde | zegt |
| streven naar | streven | streeft | gestreefd | streef | waar |
| verlangen naar | verlangen | verlangt | verlang | verlangd | waar |
| vragen naar | vragen | vroege | gevraagde | vraagt | vraag |
| bekommeren | bekommeren | bekommert | bekommerde | bekommerden | kinderen |
| om | | | | | |
| geven om | geeft | geven | gaf | geef | mensen |
| lachen om | lachen | wel | lacht | gelachen | moest |
| malen om | malen | maalt | maalde | maal | niemand |
| rouwen om | rouwt | rouwen | dood | rouw | rouwde |
| smeken om | smeekte | smeken | smeekten | hulp | gesmeekt |
| treuren om | treuren | dood | zullen | lange | wel |
| verzoeken om | verzocht | verzoeken | verzoekt | nederlands | uitlevering |
| vragen om | vraagt | vragen | gevraagde | vroege | vroegen |
| antwoorden | antwoorden | vraag | antwoordde | vragen | antwoordt |
| op | | | | | |
| azen op | aast | azen | aasde | revanche | nieuws |
| baseren op | gebaseerd | baseert | baseerde | baseren | volgen |
| beroepen op | beroepen | beroept | beriep | beriepen | volgen |
| concentreren | concentreren | volledige | moeten | zegt | eerst |
| op | | | | | |
| doelen op | doelt | doelde | doelend | daarmee | zegt |
| lijken op | lijkt | lijken | leek | wel | sterke |
| verhalen op | verhalen | kosten | schades | verhaalde | gemeentes |
| vertrouwen op | vertrouwen | vertrouwt | vertrouw | vertrouwde | wel |
| wijzen op | wijst | wijzen | wees | gewezen | wel |
| beraden over | beraden | acties | toekomst | stappen | volgens |
| beslissen over | besliste | beslissen | ander | lot | beslisten |
| discussieren | discussieren | gediscussieerd | discussieert | wel | discussieerden |
| over | | | | | |
| klagen over | klagen | klaagt | klaagde | geklaagd | klaagden |
| onderhandelen | onderhandelen | onderhandelde | onderhandelt | nieuws | contract |
| over | | | | | |
| ontfermen | ontfermen | jaar | gaat | vrouw | twee |
| over | | | | | |
| oordelen over | oordelen | oordeelt | oordeelde | rechters | geoordeeld |
| piekeren over | piekeren | piekert | pieker | piekerde | langs |
| waken over | waken | moeten | onze | blijven | veiligheid |
| zwijgen over | zwijgen | zwijgt | gezwegen | zwijg | zweeg |
| beschermen | moeten | beschermen | zichzelf | onze | mensen |
| tegen | | | | | |
| keren tegen | keren | keert | keerde | gekeerd | keerden |
| kunnen tegen | stootjes | goed | wel | verlies | waar |
| pleiten tegen | pleit | pleiten | pleitte | gepleit | vooral |

| | | | | | |
|----------------------|-------------|-------------|-----------|-------------|---------------|
| protesteren tegen | protesteren | nieuws | mensen | bewoner | gisteren |
| strijden tegen | strijden | strijdt | streed | elkaar | gestreden |
| vechten tegen | vechten | vecht | vocht | gevochten | degradatie |
| verweren tegen | verweren | verweerde | verweert | kritiek | moeten |
| verzetten tegen | verzet | verzetten | verzette | nieuws | volgen |
| wapenen tegen | wapenen | gewapend | wapent | beter | moeten |
| behoren tot | behoort | behoren | behoord | behoorden | mogelijkheden |
| beperken tot | beperken | minimum | aantal | moeten | wel |
| besluiten tot | besloten | besloot | besluiten | besluit | jaar |
| bijdragen tot | bijdragen | dragen | draagt | bijgedragen | bijdraagt |
| dwingen tot | dwingen | dwong | gedwongen | dwingt | dwongen |
| herleiden tot | herleid | herleiden | herleidt | tien | twee |
| leiden tot | leidt | leiden | leidde | geleide | volgen |
| spreken tot | spreekt | spreekt | spreken | sprak | vooral |
| verlagen tot | verlaagd | verlagen | verlaagt | procent | euro |
| verleiden tot | laten | verleiden | liet | laat | verleide |
| betichten van | beticht | betichten | betichtte | man | twee |
| dromen van | droomt | dromen | droom | droomde | waar |
| genieten van | genieten | genieter | genoten | genoot | wel |
| getuigen van | getuigt | getuigen | getuigde | respect | weinig |
| ontdoen van | ontdoen | indruk | wel | alle | zegt |
| overtuigen van | overtuigen | proberen | mensen | moeten | probeerde |
| schrikken van | geschrokken | schrok | schrikt | schrikken | schrik |
| verschillen van | verschilt | verschillen | mening | elkaars | verschillend |
| verzekeren van | verzekerde | verzekeren | verzekert | plaats | behoud |
| voorzien van | voorzien | nieuws | alle | twee | werden |
| bedanken voor | bedanken | steun | jaar | hen | inzet |
| behoeden voor | behoeden | behoedde | behoed | behoedt | doelman |
| ijveren voor | ijvert | ijveren | ijverde | geijverd | zegt |
| kiezen voor | koos | kiezen | kiest | gekozene | kozen |
| lenen voor | leent | goed | lenen | uitstekend | leende |
| pleiten voor | pleit | pleiten | pleitte | gepleit | nieuws |
| voelen voor | voelt | voelen | voelde | weinige | voel |
| vrezen voor | vrezen | vreest | gevreesde | vreesde | vrees |
| waarschuwen voor | waarschuwen | nieuws | moeten | gevaren | waar |
| zorgen voor | zorgen | zorgde | zorgt | zorgden | gezorgd |

Appendix D: Translations of all predicates

The table below contains the English and German translations of all predicates used in the crosslinguistic study (Chapter 4). The final column of the table contains the preposition (and, if applicable, case) lists used to calculate Hamming distances between all pairs of predicates.

| # | Dutch | English | German | Hamming distance lists [NL, EN, DE] |
|----|----------------------|----------------------|--|-------------------------------------|
| 1 | Danken aan | Owe to | Verdanken + <i>dat</i> | [aan, to, dat] |
| 2 | (Zich) hechten aan | Be(come) attached to | Sein Herz hängen an + <i>dat</i> | [aan, to, an-dat] |
| 3 | Liggen aan | Depend on | Liegen an + <i>dat</i> / liegen bei + <i>dat</i> | [aan, on, an-dat/bei] |
| 4 | Lijden aan | Suffer from | Leiden + <i>dat</i> | [aan, from, an-dat] |
| 5 | Zich onderwerpen aan | Submit to | Sich unterwerfen + <i>dat</i> | [aan, to, dat] |
| 6 | Zich storen aan | Mind | Sich stören an + <i>dat</i> | [aan, 0, an-dat] |
| 7 | Verbinden aan | Connect to | Verbinden + <i>dat</i> | [aan, to, mit] |
| 8 | Voldoen aan | Meet | Erfüllen + <i>acc</i> | [aan, 0, acc] |
| 9 | Zich wagen aan | Hazard | Sich heranwagen + <i>acc</i> | [aan, 0, an-acc] |
| 10 | Wijten aan | Blame on | Zuschreiben + <i>dat</i> | [aan, on, dat] |
| 11 | Berusten in | Resign (oneself) to | Sich fügen in + <i>acc</i> | [in, to, in-acc] |
| 12 | Betrekken in | Involve in | Einbeziehen + <i>acc</i> | [in, in, in-acc] |
| 13 | Delen in | Share in | Beteiligt sein an + <i>dat</i> | [in, in, an-dat] |
| 14 | Geloven in | Believe in | Glauben an + <i>acc</i> | [in, in, an-acc] |
| 15 | Zich schikken in | Conform to | Sich fügen + <i>acc</i> | [in, to, in-acc] |
| 16 | Slagen in | Succeed | Schaffen + <i>acc</i> | [in, in, acc] |
| 17 | Zich verdiepen in | Bure oneself in | Sich vertiefen + <i>acc</i> | [in, in, in-acc] |
| 18 | Zich verheugen in | Enjoy | Sich freuen + <i>acc</i> | [in, 0, über-acc] |
| 19 | Volharden in | Persist in | Durchhalten + <i>acc</i> | [in, in, acc] |
| 20 | Voorzien in | See to | Versorgen mit + <i>dat</i> | [in, to, mit] |

| | | | | |
|----|----------------------|------------------------------------|-------------------------------------|----------------------|
| 21 | Gissen naar | Guess at | Mutmaßen + <i>acc</i> | [naar, at, über-acc] |
| 22 | Luisteren naar | Listen to | Hören auf + <i>acc</i> | [naar, to, auf-acc] |
| 23 | Refereren naar | Refer to | Verweisen + <i>acc</i> | [naar, to, auf-acc] |
| 24 | Smachten naar | Pine for | Schmachten nach + <i>dat</i> | [naar, for, nach] |
| 25 | Smaken naar | Taste of | Schmecken nach + <i>dat</i> | [naar, of, nach] |
| 26 | Snakken naar | Yearn for | Sich sehnen nach + <i>dat</i> | [naar, for, nach] |
| 27 | Streven naar | Strive for | Streben nach + <i>dat</i> | [naar, for, nach] |
| 28 | Verlangen naar | Long for | Verlangen nach + <i>dat</i> | [naar, for, nach] |
| 29 | Vragen naar | Ask for | Fragen nach + <i>dat</i> | [naar, for, nach] |
| 30 | Zoeken naar | Search for | Suchen nach + <i>dat</i> | [naar, for, nach] |
| 31 | Zich bekommeren om | Worry about / concern oneself with | Sich kümmern um + <i>acc</i> | [om, about/with, um] |
| 32 | Geven om | Care about | Mögen + <i>acc</i> | [om, about, acc] |
| 33 | Lachen om | Laugh at | Lachen + <i>acc</i> | [om, at, über-acc] |
| 34 | Malen om | Care about | Such kümmern + <i>acc</i> | [om, about, acc] |
| 35 | Rouwen om | Mourn for | Trauern um + <i>acc</i> | [om, for, um] |
| 36 | Smeken om | Beg for | Flehen um + <i>acc</i> | [om, for, um] |
| 37 | Treuren om | Grieve for | Trauern um + <i>acc</i> | [om, for, um] |
| 38 | Verzoeken om | Request | Bitten um + <i>acc</i> | [om, 0, um] |
| 39 | Vragen om | Ask for | Herausfordern + <i>acc</i> | [om, for, acc] |
| 40 | Antwoorden op | Answer | Antworten auf + <i>acc</i> | [op, 0, auf-acc] |
| 41 | Azen op | Have one's eye on | Lauern auf + <i>acc</i> | [op, on, auf-acc] |
| 42 | Zich/iets baseren op | Base on | Gründen auf + <i>acc</i> | [op, on, auf-acc] |
| 43 | Zich concentreren op | Concentrate on | Sich konzentrieren auf + <i>acc</i> | [op, on, auf-acc] |
| 44 | Doelen op | Refer to | Zielen auf + <i>acc</i> | [op, to, auf-acc] |
| 45 | Lijken op | Resemble | Gleichen + <i>dat</i> | [op, , dat] |
| 46 | Zich beroepen op | Appeal to | Sich berufen + <i>acc</i> | [op, to, auf-acc] |
| 47 | Verhalen op | Recover from | Such schadlos halten + <i>dat</i> | [op, from, an-dat] |
| 48 | Vertrouwen op | Trust / rely on | Sich verlassen auf + <i>acc</i> | [op, 0/on, auf-acc] |
| 49 | Wijzen op | Indicate / point to | Hinweisen + <i>acc</i> | [op, 0/to, auf-acc] |
| 50 | Zich beraden over | Deliberate | Nachdenken über + <i>acc</i> | [over, on, über-acc] |

| | | | | |
|----|----------------------|-------------------------|-----------------------------------|-------------------------|
| 51 | Beslissen over | Decide on | Entscheiden über + <i>acc</i> | [over, on, über-acc] |
| 52 | Discussiëren over | Discuss | Diskutieren über + <i>acc</i> | [over, 0, über-acc] |
| 53 | Klagen over | Complain about | Sich beklagen über + <i>acc</i> | [over, about, über-acc] |
| 54 | Onderhandelen over | Negotiate about | Verhandeln über + <i>acc</i> | [over, about, über-acc] |
| 55 | Zich ontfermen over | Take care of | Sich erbarmen + <i>gen</i> | [over, of, gen] |
| 56 | Oordelen over | Judge | Urteilen über + <i>acc</i> | [over, 0, über-acc] |
| 57 | Piekeren over | Worry about | Grübeln über + <i>acc</i> | [over, about, über-acc] |
| 58 | Waken over | Watch over | Wachen über + <i>acc</i> | [over, over, über-acc] |
| 59 | Zwijgen over | Keep silent about | Schweigen über + <i>acc</i> | [over, about, über-acc] |
| 60 | Beschermen tegen | Protect against | Schützen gegen + <i>acc</i> | [tegen, against, gegen] |
| 61 | Zich keren tegen | Turn against | Wich wenden gegen + <i>acc</i> | [tegen, against, gegen] |
| 62 | Kunnen tegen | Can stand | Vertragen + <i>acc</i> | [tegen, 0, acc] |
| 63 | Pleiten tegen | Plead against | Plädieren gegen + <i>acc</i> | [tegen, against, gegen] |
| 64 | Protesteren tegen | Protest against | Protestieren gegen + <i>acc</i> | [tegen, against, gegen] |
| 65 | Strijden tegen | Fight against | Streiten gegen + <i>acc</i> | [tegen, against, gegen] |
| 66 | Vechten tegen | Fight against | Kämpfen | [tegen, against, gegen] |
| 67 | Zich verweren tegen | Defend oneself against | Sich wehren gegen + <i>acc</i> | [tegen, against, gegen] |
| 68 | Zich verzetten tegen | Resist | Sich widersetzen + <i>dat</i> | [tegen, 0, dat] |
| 69 | Zich wapenen tegen | Arm against | Wich sappnen gegen + <i>acc</i> | [tegen, against, gegen] |
| 70 | Behoren tot | Belong to | Gehören zu + <i>dat</i> | [tot, to, zu] |
| 71 | Zich beperken tot | Limit oneself to | Sich beschränken auf + <i>acc</i> | [tot, to, auf-acc] |
| 72 | Besluiten tot | Decide | Sich entschließen zu + <i>dat</i> | [tot, on/upon, zu] |
| 73 | Bijdragen tot | Contribute to | Beitragen zu + <i>dat</i> | [tot, to, zu] |
| 74 | Dwingen tot | Force | Zwingen + <i>dat</i> | [tot, 0, zu] |
| 75 | Herleiden tot | Reduce to | Zurückführen + <i>acc</i> | [tot, to, auf-acc] |
| 76 | Leiden tot | Lead to | Führen zu + <i>dat</i> | [tot, to, zu] |
| 77 | Spreken tot | Speak to | Sprechen zu + <i>dat</i> | [tot, to, zu] |
| 78 | Zich verlagen tot | Stoop to | Sich herablassen zu + <i>dat</i> | [tot, to, zu] |
| 79 | Verleiden tot | Tempt to | Verleiten zu + <i>dat</i> | [tot, to, zu] |
| 80 | Betichten van | Accuse of / charge with | Bezichtigen + <i>gen</i> | [van, of/with, gen] |
| 81 | Dromen van | Dream of | Träumen von + <i>dat</i> | [van, of, von] |

| | | | | |
|----|-----------------------|------------------------------|-------------------------------------|------------------------|
| 82 | Genieten van | Enjoy | Genießen + <i>acc</i> | [van, 0, acc] |
| 83 | Getuigen van | Show | Zeugen von + <i>dat</i> | [van, 0, von] |
| 84 | (Zich) ontdoen van | Strip of / dispose of | (Sich) entledigen + <i>gen</i> | [van, of, gen] |
| 85 | Overtuigen van | Convince of | (Sich) überzeugen von + <i>dat</i> | [van, of, von] |
| 86 | Schrikken van | Get scared by | Erschrecken vor + <i>dat</i> | [van, by, vor-dat] |
| 87 | Verschillen van | Differ from | Sich unterscheiden von + <i>dat</i> | [van, from, von] |
| 88 | (Zich) verzekeren van | Ensure / assure (oneself) of | (Sich) versichern + <i>gen</i> | [van, 0/of, gen] |
| 89 | Voorzien van | Provide with | Versehen mit + <i>dat</i> | [van, with, mit] |
| 90 | Bedanken voor | Decline / thank for | Bedanken für + <i>acc</i> | [voor, 0/for, für] |
| 91 | Behoeden voor | Guard from | Beschützen vor + <i>dat</i> | [voor, from, vor-dat] |
| 92 | Ijveren voor | Work for | Sich einsetzen für + <i>acc</i> | [voor, for, für] |
| 93 | Kiezen voor | Choose | Wählen + <i>acc</i> | [voor, 0, acc] |
| 94 | Zich lenen voor | Be suitable for | Sich eignen + <i>acc</i> | [voor, for, für] |
| 95 | Pleiten voor | Plead for | Plädieren für + <i>acc</i> | [voor, for, für] |
| 96 | Voelen voor | Fancy | Sich angezogen fühlen + <i>dat</i> | [voor, 0, von] |
| 97 | Vrezen voor | Fear for | Such fürchten um + <i>acc</i> | [voor, for, um] |
| 98 | Waarschuwen voor | Warn about | Warnen vor + <i>dat</i> | [voor, about, vor-dat] |
| 99 | Zorgen voor | See to / lead to | Sorgen für + <i>acc</i> | [voor, to, für] |

Appendix E: Intra preposition group dissimilarity score

Based on the predicate pair dissimilarity scores from the cross-linguistic study, a single *intra preposition group dissimilarity score* was calculated for each predicate. This score is the mean of the of all scores that compared a given predicate to others in its preposition group. The table below lists the scores for all 99 predicates. The final column, containing the score is colour marked, with green indicating lower dissimilarity and orange higher.

| P | # | Predicaat | IntraP sim |
|---|----|----------------------|------------|
| 1 | 2 | (zich) hechten aan | 1.20 |
| 1 | 1 | Danken aan | 1.30 |
| 1 | 5 | Zich onderwerpen aan | 1.30 |
| 1 | 6 | Zich storen aan | 1.30 |
| 1 | 9 | Zich wagen aan | 1.43 |
| 1 | 4 | Lijden aan | 1.50 |
| 1 | 7 | Verbinden aan | 1.50 |
| 1 | 10 | Wijten aan | 1.50 |
| 1 | 3 | Liggen aan | 1.53 |
| 1 | 8 | Voldoen aan | 1.60 |
| 2 | 12 | Betrekken in | 1.00 |
| 2 | 17 | Zich verdiepen in | 1.00 |
| 2 | 16 | Slagen in | 1.20 |
| 2 | 19 | Volharden in | 1.20 |
| 2 | 13 | Delen in | 1.25 |
| 2 | 14 | Geloven in | 1.25 |
| 2 | 11 | Berusten in | 1.30 |
| 2 | 15 | Zich schikken in | 1.30 |
| 2 | 20 | Voorzien in | 1.60 |
| 2 | 18 | Zich verheugen in | 1.80 |
| 3 | 24 | Smachten naar | 0.70 |
| 3 | 26 | Snakken naar | 0.70 |
| 3 | 27 | Streven naar | 0.70 |
| 3 | 28 | Verlangen naar | 0.70 |
| 3 | 29 | Vragen naar | 0.70 |
| 3 | 30 | Zoeken naar | 0.70 |
| 3 | 25 | Smaken naar | 1.20 |
| 3 | 22 | Luisteren naar | 1.60 |
| 3 | 23 | Refereren naar | 1.60 |
| 3 | 21 | Gissen naar | 1.80 |
| 4 | 35 | Rouwen om | 1.00 |
| 4 | 36 | Smeken om | 1.00 |
| 4 | 37 | Treuren om | 1.00 |
| 4 | 31 | Zich bekommeren om | 1.22 |
| 4 | 39 | Vragen om | 1.22 |

| | | | |
|---|----|-----------------------|------|
| 4 | 38 | Verzoeken om | 1.33 |
| 4 | 32 | Geven om | 1.39 |
| 4 | 34 | Malen om | 1.39 |
| 4 | 33 | Lachen om | 1.78 |
| 5 | 41 | Azen op | 0.85 |
| 5 | 42 | Zich/iets baseren op | 0.85 |
| 5 | 43 | Zich concentreren op | 0.85 |
| 5 | 48 | Vertrouwen op | 0.88 |
| 5 | 49 | Wijzen op | 0.93 |
| 5 | 44 | Doelen op | 0.95 |
| 5 | 46 | Zich beroepen op | 0.95 |
| 5 | 40 | Antwoorden op | 1.00 |
| 5 | 45 | Lijken op | 1.80 |
| 5 | 47 | Verhalen op | 1.80 |
| 6 | 53 | Klagen over | 0.70 |
| 6 | 54 | Onderhandelen over | 0.70 |
| 6 | 57 | Piekeren over | 0.70 |
| 6 | 59 | Zwijgen over | 0.70 |
| 6 | 50 | Zich beraden over | 0.90 |
| 6 | 51 | Beslissen over | 0.90 |
| 6 | 52 | Discussiëren over | 0.90 |
| 6 | 56 | Oordelen over | 0.90 |
| 6 | 58 | Waken over | 1.00 |
| 6 | 55 | Zich ontfermen over | 1.80 |
| 7 | 60 | Beschermen tegen | 0.40 |
| 7 | 61 | Zich keren tegen | 0.40 |
| 7 | 63 | Pleiten tegen | 0.40 |
| 7 | 64 | Protesteren tegen | 0.40 |
| 7 | 65 | Strijden tegen | 0.40 |
| 7 | 66 | Vechten tegen | 0.40 |
| 7 | 67 | Zich verweren tegen | 0.40 |
| 7 | 69 | Zich wapenen tegen | 0.40 |
| 7 | 62 | Kunnen tegen | 1.70 |
| 7 | 68 | Zich verzetten tegen | 1.70 |
| 8 | 70 | Behoren tot | 0.40 |
| 8 | 73 | Bijdragen tot | 0.40 |
| 8 | 76 | Leiden tot | 0.40 |
| 8 | 77 | Spreken tot | 0.40 |
| 8 | 78 | Zich verlagen tot | 0.40 |
| 8 | 79 | Verleiden tot | 0.40 |
| 8 | 71 | Zich beperken tot | 1.00 |
| 8 | 75 | Herleiden tot | 1.00 |
| 8 | 72 | Besluiten tot | 1.10 |
| 8 | 74 | Dwingen tot | 1.10 |
| 9 | 81 | Dromen van | 1.20 |
| 9 | 85 | Overtuigen van | 1.20 |
| 9 | 84 | (zich) ontdoen van | 1.30 |
| 9 | 88 | (zich) verzekeren van | 1.33 |
| 9 | 83 | Getuigen van | 1.35 |

| | | | |
|----|----|------------------|------|
| 9 | 80 | Betichten van | 1.38 |
| 9 | 87 | Verschillen van | 1.50 |
| 9 | 82 | Genieten van | 1.65 |
| 9 | 89 | Voorzien van | 1.75 |
| 9 | 86 | Schrikken van | 1.80 |
| 10 | 92 | Ijveren voor | 1.05 |
| 10 | 94 | Zich lenen voor | 1.05 |
| 10 | 95 | Pleiten voor | 1.05 |
| 10 | 90 | Bedanken voor | 1.10 |
| 10 | 99 | Zorgen voor | 1.40 |
| 10 | 97 | Vrezen voor | 1.45 |
| 10 | 93 | Kiezen voor | 1.65 |
| 10 | 96 | Voelen voor | 1.65 |
| 10 | 91 | Behoeden voor | 1.70 |
| 10 | 98 | Waarschuwen voor | 1.70 |