# HOW, WHY AND WHERE

## DOES ARGUMENT STRUCTURE VARY?

A usage-based investigation into the Dutch transitive-prepositional alternation

Dirk Pijpops

# How, why and where does argument structure vary?
---
# A usage-based investigation into the Dutch transitive-prepositional alternation

Thesis presented in partial fulfillment of the requirements for the degree of Doctor in Linguistics

by Dirk Pijpops

Supervisor:          Dirk Speelman
Co-supervisors:   Freek Van de Velde
                          Stefan Grondelaers

Leuven, 2019

*Almost, at times,*


*the Fool.*

-    The Love Song of J. Alfred
     Prufrock, Thomas Eliot

Aan Peggy

# Table of contents

# Dankwoord

Allereerst gaat mijn dank uit naar mijn promotoren Dirk Speelman, Freek Van de Velde en Stefan Grondelaers. Dirk, ik dank je niet alleen voor je jarenlange begeleiding, maar ook omwille van je inspirerende gedrevenheid om diepgravend wetenschappelijk werk te verrichten. Ook als ik de weg niet meer zag, behield jij je kalme zekerheid dat het goed kwam. Freek, het valt me moeilijk op te sommen waar ik je het meest dankbaar voor ben. Hier een poging: ik kan steeds op je rekenen voor professionele en persoonlijke raad – en je staat maar al te graag open voor discussie als ik het niet (onmiddellijk) eens ben met die raad; je bent een onuitputtelijke bron van ideeën; en je bent altijd bereid je benijdenswaardig diepe inzicht in het Nederlandse taalsysteem te delen. Stef, je raadde me aan diep op mijn alternantie in te zoomen, en stimuleerde me vervolgens om ook weer uit te zoomen. Dat advies loopt als een rode lijn door deze dissertatie. Bovendien zette je me aan steeds kritisch te blijven en helder te formuleren, waarvoor mijn dank.

Naast mijn promotoren dienen de overige leden van mijn begeleidings-commissie en jury hier zeker vermeld te worden. Aan Kris Heylen ben ik dank verschuldigd vanwege zijn bereidheid mij steeds opnieuw bij te staan, onder andere bij het aanmaken van en rekenen met distributionele vectoren, maar ook met zijn sterke theoretische doorzicht. Mijn dank gaat ook uit naar Roeland van Hout, die me telkens een nieuw perspectief wist te bieden op mijn onderzoek. Liesbeth Augustinus heeft bij het begin van dit onderzoek veel tijd en moeite uitgetrokken om een jonge doctoraatstudent uit een andere onderzoeksgroep op weg te zetten bij het gebruik van het Sonar-corpus, waarvoor mijn welgemeende dank. Bovendien vormde haar eigen dissertatie een inspiratiebron voor dit werk. Evenzo ben ik Timothy Colleman en Florent Perek dankbaar om in mijn jury te zetelen. Wie bekend is met hun werk – en wie dat niet is, doet er goed aan zich dringend in te lezen – ziet er duidelijk de invloed van doorheen de hele tekst.

Daarbuiten gaat mijn dank uit naar de docenten van mijn opleiding: naast diegenen die hierboven al genoemd zijn, wisten vooral Dirk Geeraerts, Luc Draye en Joop van der Horst mijn liefde voor de taalkunde aan te wakkeren met hun interessante en vermakelijke lessen. Ook ben ik erkentelijk aan het Fonds Wetenschappelijk Onderzoek om mogelijk te maken dat ik me vier jaar kon kwijten van taalkundig onderzoek, alsook de KU Leuven, mijn

thuisuniversiteit, onder andere om mijn onderzoekverblijf in Nijmegen te financieren.

Ik ben steeds met veel plezier naar kantoor gegaan. Dat had te maken met het werk zelf, maar nog veel meer met de collega's, in Brussel, Leuven en Nijmegen. Ik dank Paul, Remi, Pieter, Yana, Luc, Johan en Miquel, die me onmiddellijk in hun midden welkom hebben geheten en me kansen hebben geboden waar ik enkel van kon dromen. Katrien, dat laatste geldt in de overtreffende trap voor jou. Jouw voorbeeld en begeleiding hebben me geïnspireerd het beste van mezelf te geven.

Ik richt ook een woord van dank tot mijn kantoorgenoten, voor zover die nog niet aan bod zijn gekomen: tot Benedikt, voor onze amusante gesprekken over misvattingen in de statistiek en andere dwalingen in ons veld, en tot Robbert, die een inmiddels oude knar vaak wist op te beuren met zijn aanstekelijke enthousiasme. Natuurlijk dank ik daarnaast Laura, Lauren, Eline, Jocelyne, Jason, Melanie, Kristina, Lena, Leonie, Jeroen, Stefano, Alek, Isabeau, Danqing, Mariana, Alexandra, Jozefien, Tanguy en alle anderen voor de fijne koffies, lunches, occasionele diners en zo meer. Mijn innig gemeende excuses voor het gevloek dat bijwijlen door de gang klonk: dat was steeds gericht op mijn code, nooit op jullie. Karlien, dank je voor je steun wanneer het tegenzat, voor het vele plezier dat we gehad hebben, maar ook voor de nieuwe inzichten die je aanbracht tijdens onze uitgebreide discussies over, onder andere, taalkunde. Jouw bijdrage aan dit doctoraat is wellicht groter dan je zelf vermoedt.

Mijn dank gaat tevens uit naar de studenten de ik heb mogen begeleiden en die op hun beurt mij begeleid hebben, met name Catho Jacobs en Astrid Vandervaeren. Ik ben ook dank verschuldigd aan Antal van den Bosch, voor zijn hulp tijdens mijn verblijf in Nijmegen en daarna, en natuurlijk aan mijn Nijmeegse collega's Louis, Claire, Thijs, Theresa, Chantal, Aurora en Katherine.

Met evenveel plezier trok ik de deur van het kantoor 's avonds echter achter me dicht, en dat is dan weer in niet geringe mate te wijten aan mijn vrienden. Ik dank Henri, Brecht en Jonas, wier deskundige oordelen rechtstreeks hebben bijgedragen aan dit doctoraat, maar wier contributie toch voornamelijk te situeren is tussen pot en pint, alsook Bram, Kristof, Stef en Bruno, die me telkens uit de academische bubbel wisten te sleuren om me te tonen wat echt belangrijk is: barbecues, nonsens en vriendschap.

Tot slot wil ik mijn familie bedanken. Ten eerste dank ik mijn ouders en grootvader voor hun onvoorwaardelijke jarenlange steun, alsook mijn zussen Tine en Lize en mijn broer Jan, en natuurlijk John, Lien, Yannick, Caressia, tante Begga en de ganse familie Verboven, met speciale vermelding van Gijsieboy. Jullie gaven me een warme thuis en stonden me bij in alles wat ik

ondernam. Ten tweede richt ik me tot mijn schoonfamilie, Gerda en Eduard, Johan en Sandra, Jeroen en Ellen. Bedankt voor alle hulp en alles wat jullie me hebben geleerd – al had dat niet rechtstreeks met dit onderzoek te maken. Maar evenzeer ben ik jullie dankbaar voor alle aanmoedigingen; een fijnere schoonfamilie kan men zich niet wensen.

Mijn allerlaatste en meest innige dankwoord bewaar ik voor mijn belangrijkste inspiratiebron, mijn compagnon de route, mijn vrouw. Peggy, alles wat ik de afgelopen 11 jaar bereikt heb, hebben we samen bereikt. Zo ook de voltooiing van dit doctoraat; enkel jij en ik kunnen bevatten hoeveel hulp ik aan jou gehad heb. Weet dat ik het niet alleen door jou gedaan heb, maar ook voor jou. Voor je al je vrolijkheid, begrip, en liefde: dank je. Voor alles: dank je.

# 1     Introduction

As the Dutch language lost its case system during the deflection process that has characterized many of the major European languages in the last millennium, its verbs needed to seek new ways of organizing their argument structure (van der Horst 2008: 1620–1621). Verbs with two complements that had previously used case marking to distinguish between both, were now faced with roughly two options. The first was to consolidate around the time-honored transitive construction; the second was to join the new kid on the block, the prepositional object (Duinhoven 1989; van der Horst and Van de Velde 2008: 58).[1] Most verbs would eventually settle down on either strategy, albeit that some first went with the transitive construction and later switched teams to the prepositional object, e.g. *wachten op* 'wait', or vice versa, e.g. *trachten naar* 'attempt' (van der Horst 2008: 1857–1858). Still, particular verbs seemingly believed they could have their cake and eat it, and continued to appear at times in the transitive construction or in a construction with a subject and a prepositional object – what we will call a prepositional-intransitive construction.[2] This resulted in what appears to be free synchronic variation. The *Algemene Nederlandse Spraakkunst* or *General Dutch Grammar* puts it as follows (Haeseryn et al. 1997: 1168):

> *As for its meaning, the prepositional object exhibits resemblance to the direct object. Hence, there exist nearly or completely identical verbs that may combine both with a direct or a prepositional object without a difference in meaning. In the following sentences, the constituents in bold are direct object in the a-sentences, and prepositional object in the b-sentences.* [Our translation]

1.     a.     *Zoek je* **je paraplu**?
            Search you   your umbrella
            'Are you looking for your umbrella?'

---

[1] It is actually unclear what is cause and what is effect in these developments (cf. Dale and Lupyan 2012; van der Horst 2013; Bentz et al. 2015). This is not the concern of the present study, however.

[2] The name *prepositional-intransitive construction* is adopted from Sag, Wasow and Bender (2003: 237, 241, 496). It is defined in Chapter 3.

b. *Zoek je* **naar je** **paraplu***?*
   Search you to    your umbrella
   'Are you looking for your umbrella?'

2. a. *Hij heeft* **een rijke vrouw** *getrouwd.*
      he has    a    rich woman married
      'He has married a rich woman.'

   b. *Hij is* **met een rijke vrouw** *getrouwd.*
      he is  with a    rich woman married
      'He has married a rich woman.'

3. a. *Ik betwijfel***de betrouwbaarheid van deze proef** *in sterke mate.*
      I    doubt    the reliability    of    this    test    in strong measure
      'I seriously doubt the reliability of this test.'

   b. *Ik twijfel in sterke mate* **aan de betrouwbaarheid van deze**
      I   doubt   in strong measure on    the reliability    of    this
      **proef***.*
      test
      'I seriously doubt the reliability of this test.'

This excerpt immediately raises a number of questions. First, is there really no meaning difference whatsoever between both forms, not even a very subtle one? Are there perhaps other verbs for which the presence or absence of the same prepositions does signify a meaning difference? And if meaning is not the (only) determinant, what other factors influence the appearance of these prepositions? Is one of the variants perhaps easier to understand or to produce, or do particular language users prefer one variant over the other? Second, why do such concerns – typically called alternation factors – actually come to determine the presence of the prepositions? What has caused or is causing their influence? And third, is this in fact a single alternation at all, or are they multiple alternations that just happen to look alike? Do the same factors determine the choice between both variants, regardless of the verb, or should we look at each verb separately? Or is it even a different story for each object that the verbs combine with?

    These questions are specific cases of more general questions regarding verbal argument structure, viz. how, why and where does argument structure vary? A lot is already known about these three questions. The discipline of linguistics has made good progress in answering them, and each of them has sparked in-depth theoretical discussions and even research traditions in their own right. The goal of this thesis is to make a start of answering these questions for the transitive-

prepositional alternation – or alternations – and in doing so, we hope to show how tackling these questions together may provide more insight than dealing with them separately. We will only consider verbs that are fully identical in form, as in (1)-(2), and furthermore focus on those alternating with the preposition *naar* 'to', as in (1). Alternations such as (3) fall outside of the scope of this study. The reason is that we want to be sure that any subtle difference in meaning, language processing or lectal preference between the variants is due to the presence of the preposition, and not due to a verbal prefix such as *be-* in (3) (cf. Grondelaers, Speelman and Geeraerts 2008: 158).

The present research is set up within the framework of usage-based linguistics (Tomasello 2003; Bybee 2006; Diessel 2017). This does not mean we will ignore findings from other traditions, since a lot of answers to the how-, why- and where-questions have been achieved in frameworks that are not usage-based in a strict sense. It does mean two things, however. First, it means that we consider language use or the directly observable behavior of language producers to be our object of study (cf. Subsection 2.1. on the employed conception of 'grammar'). This behavior, as it is registered in the utterances that make up a corpus or can be examined in the context of an experiment, thus forms the primary empirical data of usage-based linguistics, rather than a native speaker's intuitions about their language.[3] Second, it means that we do not seek to de- or prescribe which sentences are grammatical and which are not. Instead, we merely aim to explain said linguistic behavior of language producers.[4]

The current chapter provides an introduction to this thesis by shortly discussing the how-, why-, and where-questions.

---

[3] Of course, these intuitions can still be useful either in drafting hypotheses or even in testing hypotheses, provided that they are gathered and analyzed with sufficient care and on a sufficiently large scale (Gilquin and Gries 2009: 3; for examples, see Verhagen and Maria 2016 and references cited therin).

[4] Arppe et al. (2010: 18–19), however, present a compelling argument why intuitions about the familiarity or 'grammaticality' of utterances should still be studied as a research object in their own right by usage-based linguistics. It goes as follows. Since language users are (subconsciously) constantly judging the utterances of their interlocutors, e.g. in order to classify these interlocutors as either 'us' or 'them', such judgments constitute a form of natural linguistic behavior – albeit behavior of a different quality than that of language production. However, we do not aim to study this type of linguistic behavior in the present thesis: we focus on the linguistic behavior of language producers.

## 1.1     How does argument structure vary?

Answering the question *how does argument structure vary?*, or more generally, *how does any particular alternation that is put under scrutiny, vary?* forms the chief objective of the research tradition of alternation studies. Such studies take an alternation of two or more seemingly interchangeable variants and try to find out what factors drive the choice between these variants (for foundational studies, see Labov 1966a; Grondelaers 2000; Gries 2003; Colleman 2006, for examples of more recent work, see Claes 2015; Krawczak and Glynn 2015; Szmrecsanyi et al. 2017). An alternation study involves gathering a number of instances of the variants, which are typically drawn from a corpus, and searching for correlations between the choice of variant and a number of independent variables.

   For instance, Grondelaers (2000) extracted a number of instances like (4)-(5) from the ConDiv-corpus of written Dutch. The variants at issue here are the variant with *er* 'there', as in (4a) and (5a) and without *er* 'there,' as in (4b) and (5b). He then searched for and found correlations between the choice of variant and a number of independent variables, like ADJUNCT TYPE with the values *locative*, as in (4), and *temporal* as in (5). That is, *er* 'there' appeared significantly more often with temporal adjuncts than with locative ones.

4.   a.   *Over het platgelopen jaagpad naderde  **er**  dan een man.*
          over   the  well-trodden  towpath  approached there then a     man
          'Over the well-trodden towpath, a man then approached'

     b.   *Over het platgelopen  jaagpad naderde     dan een man.*
          Over the well-trodden   towpath, approached then a     man
          'Over the well-trodden towpath, a man then approached'
                              (taken over from Grondelaers 2000: 172)

5.   a.   *Tijdens de   pauze naderde   **er**   dan een man.*
          During  the   break, approached there then a     man
          'During the break, a man then approached.'

     b.   *Tijdens de pauze naderde    dan een man.*
          During  the break  approached then a     man
          'During the break, a man then approached.'

In alternation studies, such observed correlations are usually taken to indicate causal relations. These relations do not necessarily hold directly between the

independent variables and the choice of variant, but are at least presumed to hold between some underlying entity that is operationalized by one or more independent variables, and the choice of variant (Stefanowitsch 2010: 370–372). For example, Grondelaers (2000: 134–136) argues that locative adjuncts generally render the following subject of the sentence fairly predictable, or at least to a greater extent than temporal adjuncts tend to do. This would result in *er* 'there' being more often dropped after locative adjuncts (for additional evidence, see Grondelaers et al. 2009).

Researchers are often specifically interested in one type of independent variable. For instance, Levin and Grafmiller (2012) mainly focus on a semantic difference between English psych verbs, while Szmrecsanyi (2013) or Geleyn (2017) are primarily interested in the time variable, and Szmrecsanyi et al. (2016) are first and foremost concerned with lectal differences, i.e. differences between several varieties of a language, notably regiolects. Still, all of these studies do take other types of variables into account, e.g. relating to language processing. The present study subscribes to this tradition.

The goal of answering the how-question is in the first place descriptive: this thesis hopes to contribute to the description of the Dutch language by describing the factors that govern the transitive-prepositional alternation in (1). We will not claim to be exhaustive in this regard, however: the present investigation is only the first in-depth corpus study of the Dutch transitive-prepositional alternation. Moreover, even if it were not the first study, exhaustibility would still likely be an overambitious goal. The English dative alternation, by comparison, has been subject to numerous studies, but still many research questions remain, e.g. regarding the influence of lexical retrieval speed (Dubois forthc., for previous book-length studies, see Oehrle 1976; Theijssen 2012; Bürkle 2015; Röthlisberger 2018a). Instead, we merely aim to provide a first exploration of the alternation, and hopefully identify the most crucial alternation factors. To find these, we will base our investigation upon the hypotheses of the second question, viz. the why-question.

## 1.2     Why does argument structure vary?

The question *why does argument structure vary?* asks for the causes of the alternation factors. If we are able to uncover them, we may discover why language variation arises and how it may often remain diachronically stable. We discern three loci from which factors may emerge that determine the alternation.

The first is the lectal context. By *lectal*, we refer to anything that has to do with differences between varieties of a language, e.g. dia-, socio-, regio- or ethnolects

(Geeraerts 2005). This also includes the broad socio-cultural context a language is set in. For example, this type of information is needed to explain the alternation between the Dutch second person singular pronouns *gij* and *jij* (Philippa et al. 2009). The standard language form is *jij*, while *gij* functions as an archaic and extremely formal variant in the North of the Dutch language area in Europe – in a way that is similar to *thou* in English – whereas it is, at the same time, the colloquial, informal variant in the South.

The origin of this seemingly contradictory arrangement is to be found in the changing social prestige of Dutch dialects during the history of Dutch. Just when the Dutch standard language started to develop, the cultural and economic center of the Low Countries shifted from the duchy of Brabant in the South to the county of Holland in the North (more on this in Section 4.1). As a result, the earliest stages of would-be standard Dutch still drew on the Brabantic dialects, but this standard was later engrafted onto the Hollandic dialects. The original form *gij* came to be pronounced as *jij* in the Hollandic dialects, and eventually replaced *gij* in the standard language, while *gij* was retained as the normal form in the Southern dialects. The same situation, viz. one form acting as the archaic, formal variant in the North and simultaneously as the more colloquial, informal variant in the South, holds for a large number of alternations in Dutch, including *wenen-huilen* 'cry with tears', *nagel-spijker* 'nail' and *nochtans-toch* 'still' (den Boon and Geeraerts 2005; Haeseryn 2013: 711). To formulate hypotheses regarding this type of alteration factor, we will turn to the field of (historical) sociolinguistics.

A second domain from which alternation factors may develop is semantics. Like the lectal context, semantics is also meant in the broad sense, including anything that has to do with meaning expression or construal, e.g. notions like affectedness, aspect, causation or concepts such as Proto-Agents and Proto-Patients, flow of energy, etc. (cf. respectively Verhagen 2007; Beavers 2011; van Hout 1996; Verhagen and Kemmer 1997; Dowty 1991; and Langacker 1991 and van Voorst 1996).

An example can be found in the choice between Dutch causative auxiliaries *doen* and *laten* (Speelman and Geeraerts 2009; Levshina, Geeraerts and Speelman 2013) or in the English conative alternation (Broccias 2001; Perek 2015). When a language user wants to construe a causative chain as involving a direct causation, he/she is more likely to use *doen*, whereas *laten* functions as the alternative for indirect causation (Levshina, Geeraerts and Speelman 2013). For the English conative alternation, multiple meaning differences appear to be at play, depending on the alternating verb. These include whether the action expressed by the verb occurred in a bit-by-bit manner, whether it was repeated, or whether it was only partially executed (Perek 2015: 111–142). To formulate hypotheses about which meaning differences may arise, we will turn to the field of cognitive linguistics, that regards the study of meaning as its core business (Geeraerts 2006a,b).

A third possible domain of origin for alternation factors is language processing, i.e. the cognitive and physical machinery that is used to encode a proposition or a point of view into an utterance, send that utterance to an interlocutor, receive it and decode it. This domain is different from the semantic domain in that the influence of the alternation factors are not caused by a meaning difference. We explicitly include physical machinery, because the physical limitations of the channels of human language clearly have some effect on language. For example, tone languages such as Chinese of course do not employ tones that the human ear cannot receive or the human larynx cannot produce. Moreover, information-theoretic accounts of the influence of complexity on language variation explicitly refer to physical constraints on the language channel (cf. Subsection 4.3.1.2).

The most typical example illustrating the influence of language processing on linguistic variation is perhaps the English *that*-alternation, as in (6). Here, it is found that the complexity of the subject of the subordinate clause for a large part determines the presence of *that* (Torres Cacoullos and Walker 2009; Shank, Bogaert and Plevoets 2016a). Of course, other factors are known to be at play as well (for an overview, cf. Shank, Bogaert and Plevoets 2016a: 32–44). To formulate predictions based on this type of hypotheses, we will draw from the literature of psycholinguistics.

6.   a.   *I think that he is a powerful man.*
     b.   *I think he is a powerful man.*
                    (COCA, cited in Shank, Bogaert and Plevoets 2016a: 32)

It could be proposed that a fourth type of alternation factor is lexical in nature, viz. lexical biases for either variant. However, we do not consider lexical biases to be a root cause in themselves. For instance, lexical biases may have a lectal origin, as in the case of lectal contamination (Pijpops and Van de Velde 2018a), or a semantic origin, as has been argued for the English and Dutch dative alternations by Gries and Stefanowitsch (2004: 104–107) and Colleman (2009a), or a processing-related origin, such as lexical accessibility (Ferreira and Dell 2000; Dubois forthc.) or token frequency (De Smet and Van de Velde 2019).

The goal of the why-question is broader than that of the how-question. Here, we mean to formulate and test hypotheses from historical sociolinguistics, cognitive linguistics and psycholinguistics to learn something more about the functioning of language in general, rather than just the alternation at hand. The Dutch transitive-prepositional alternation in effect only serves as a case study.

## 1.3     Where does argument structure vary?

The question *where does argument structure vary?* can be rephrased as *at which level of abstraction is argument structure determined?*. This is both a practical question in the context of an alternation study, and constitutes the focal point of a discussion in the theoretical literature. We can illustrate the first by looking at the Dutch psych verb alternation in (7)-(8), studied in Pijpops and Speelman (2017). The question is how to know whether (7) and (8) are actually examples of a single alternation, and not two separate alternations. That is, perhaps (7) and (8) do not have anything to do with each other.

There are at least two ways of looking at this alternation. The first way is to say that it is indeed a single alternation between two fairly schematic argument constructions, that simply happen to combine with different verbs in (7) and (8). If so, we should take up instances of both *ergeren* 'annoy' and *interesseren* 'interest' in a single dataset. The second way is to view (7) as one alternation between the more concrete transitive *ergeren*-construction and the reflexive *ergeren*-construction, and (8) as another alternation between the transitive *interesseren*-construction and the reflexive *interesseren*-construction. In that case, taking up (7) and (8) in the same dataset would be akin to throwing instances of the English dative and locative alternations into the same pool.

7.    a.    *Elizabeth ergert   John.*
            Elizabeth  annoys  John
            'Elizabeth annoys John.'
                        (Transitive construction, stimulus-subject; taken over from
                        Pijpops and Speelman 2017: 210–211)

      b.    *John ergert   zich    aan  Elizabeth.*
            John annoys  himself  on    Elizabeth
            'Elizabeth annoys John.'
                        (Reflexive construction, experiencer-subject)

8.    a.    *Oh ja, dat  interesseert veel   mensen.*
            Oh yes that  interests     many  people
            'Oh yes, a lot of people find that interesting.'
                        (Transitive construction, stimulus-subject; taken over from
                        (Pijpops and Speelman 2017: 230)

> b. *Oh ja, veel mensen interesseren zich daarvoor*
> oh yes many people interest themselves therefor
> 'Oh yes, a lot of people find that interesting.'
>
> (Reflexive construction, experiencer-subject)

Meanwhile, the question *at which level of abstraction is argument structure determined?* is also relevant from a theoretical point of view. The theoretical literature presents two big contenders for this title, viz. the individual verbs and more schematic argument structure constructions (see the overview in Perek 2015: 15–27).[5] On the one hand, the verb-position is taken among others by lexical approaches to argument structure, according to which the argument structure of a verb is specified in its lexical entry (Müller and Wechsler 2014; Müller 2018). If a verb then exhibits multiple argument structures, these either all need to be specified for the verb, or they may be handled by lexical rules. Lexical rules are typically meaningless, and so cannot account for consistent meaning differences, and whether or not they can apply often still need to be specified in the lexical entries of the individual verbs (Meurers and Minne 1997; van Trijp 2015: 619).

On the other hand, schematic argument structure constructions essentially constitute argument structures in their own right that are posited to exist independently of verbs and are associated with their own constructional meaning (Goldberg 1995).[6] They may in principle combine freely with verbs, although they can of course have (extremely) outspoken preferences as to with which verbs they combine in actual practice (Goldberg 2006: 22, 2013: 439–440). For instance, what is often cited as a crucial condition for argument structures and verbs to combine, is sufficient coherence between the meaning of the argument construction and the meaning of the verb. This is called the Principle of Semantic Coherence (Goldberg 1995: 50–52, 2006: 39–40, Section 4.2). The argument constructions proposed in these accounts typically rank fairly highly in terms of abstraction from specific occurrences (Goldberg 2003: 219–220, 2013: 436), although they are not necessarily fully abstract (Perek 2015: 105–111, 217).

Various researchers have argued that the discussion regarding the role of individual verbs versus schematic argument constructions should be treated as an empirical question to which the answer may very well depend on the language or even the case study at issue (Croft 2001: 28; Boas 2014; Pedersen 2016, 2019). If

---

[5] *Argument structure constructions* are called *argument constructions* in the remainder of this thesis, for reasons of brevity.

[6] There is another discussion ongoing on the question whether all constructions – and by extension all argument constructions – are necessarily meaningful, i.e. whether they all contribute some meaning to utterances, independently of the other constructions that they combine with (see the overview in Hilpert 2014a: 50–57). We will only deal with meaningful argument constructions in this thesis, however, and do not discuss this issue any further.

that is the case, however, we will need an empirical criterion on the basis of which the matter can be decided. Grammaticality is often used to this end (Goldberg 1995: 9–18; Müller and Wechsler 2014: 15–44 and references cited therein), but this notion is problematic in usage-based linguistics (for in-depth critiques of the notion of grammaticality in linguistic research, see Stefanowitsch 2007; Sampson and Babarczy 2014; van Trijp 2015: 619–621). Moreover, most approaches to argument structure that employ this criterion have thus far seemed to run circles between over- and undergeneration (van Trijp 2015: 615–619). Instead, van Trijp (2015) and Perek (2015) propose to evaluate both approaches as to how successful they are at accounting for actual language usage:

> [T]he question of what level of generalization best reflects speakers' knowledge of constructions (at least as it pertains to their semantic contribution) can be decided on the basis of usage data. (Perek 2015: 218)

Of course, we then first need to know how language is actually used, and most notably, at what level of abstraction argument structures express meaning differences. For example, the Dutch psych verb alternation in (7)-(8) appears to express a fairly consistent distinction in agentivity across multiple verbs: the variant in (7a) and (8a) seems to construe the mental event – in this case, respectively the annoyance and the interest – as involving a more agentive stimulus than the variant in (7b) and (8b) (Pijpops 2013; Pijpops and Speelman 2017). The stimulus is the participant causing the mental event, e.g. *Elizabeth* in (7) or *dat/daarvoor* 'that/therefor' in (8). As such, it may be argued that the meaning difference in (7)-(8) functions at a level of abstraction higher than that of the individual verbs, and that the alternation is also determined at this higher level of abstraction.

A similar case has been made for the dative alternation. Here, the semantic distinction between 'transfer of possession' or 'caused reception' versus 'material transfer' or 'caused motion' has proven capable of capturing the relevant semantic contrast for a large number of alternating verbs (Wierzbicka 1988; Goldberg 1992; Van Belle and Van Langendonck 1996; Colleman and De Clerck 2009). Such findings can be used as arguments to claim that the meaning differences expressed by differing argument structures need to be specified at a fairly high level of abstraction (Goldberg 2013).

Still, several researchers have noted problems with this argument, at least for the English dative alternation, as in (9). For one, Gries and Stefanowitsch (2004: 107) note that a number of commercial transaction verbs, such as *sell, supply* and *pay,* appear to defy the general semantic contrast by preferring the prepositional dative even though they first and foremost entail a transfer of possession, rather

than a material transfer. Colleman (2009a: 609) describes how such verbs can be brought back into line, however. For another, Boas (2010) and Röthlisberger, Grafmiller and Szmrecsanyi (2017) argue that meaning differences that are specific to particular lexical items may have been swept under the rug in the study of the dative alternation.

This last point actually holds more generally for usage-based studies investigating argument constructions: the level of abstraction at which these studies tend to focus, leans more toward the abstract side of the cline (Perek 2015: 105). This introduces a potential bias in the literature in favor of accounts that employ highly schematic argument constructions. There are at least two reasons for such a bias; the first of which is the high token frequency of schematic constructions compared to more concrete constructions. For instance, even in smaller corpora of English such as the ICE-GB subcorpus used by Gries and Stefanowitsch (2004), it is still easy to find hundreds of instances of the fully abstract ditransitive and prepositional dative constructions. However, it could also be conjectured that the dative alternation should actually be studied separately for each verb, such as *send, sell, pay*, etc. In that case, we would want to gather a separate dataset for e.g. the verb *pay* and treat the dative alternation with *pay* as a separate alternation in its own right. It would be a lot harder, however, to harness sufficient instances for such an alternation study.

9.  a.   *John gave/sent/brought/sold/… Mary the book.*
    b.   *John gave/sent/brought/sold/… the book to Mary.*

The second reason is simply that there are many more alternations between concrete, verb-specific constructions that would need to be to investigated than between fully schematic constructions. As such, zooming in on more concrete constructions necessarily diminishes the potential descriptive yield of an alternation study – at least at first sight: it is much nicer to discover new determinants of thé English dative alternation, than to discover new determinants of the dative alternation that are only at play for the verb *pay*.

The first reason is a merely practical impediment for the study of more concrete constructions, and one that is easily resolved as corpora become bigger. The verb *pay*, for instance, already yields 1,921,576 instances in the 7.9 billion word NOW corpus (Davies 2013), which is more than sufficient to be subjected to alternation study in its own right.[7] In fact, researchers have focused on particular lexemes in

---

[7] Of course, not all of these instances appear in the ditransitive or the prepositional dative construction, but we suspect that even after weeding out non-alternating instances, a sufficient number of instances will be left. The NOW corpus is updated daily to contain more

the study of the dative alternation, most notably the highly frequent verb *give* (Bresnan and Hay 2008; Bernaisch, Gries and Mukherjee 2014), but also *offer* and *show* (Lehmann and Schneider 2012: 72–74). Lehmann and Schneider (2012) take this approach even further, and investigate the dative alternation for instances where the verb, direct object and prepositional object slots are kept constant to particular types, e.g. *pay it attention* vs. *pay attention to it*. Of course findings based on a single verb or a single verb-theme-recipient combination do not necessarily generalize to more abstract levels, which is the point of the second reason (Röthlisberger, Grafmiller and Szmrecsanyi 2017: 675).

This second reason is based on a misapprehension, however, albeit an understandable one. Just because research findings do not have a broad descriptive coverage, does not mean that they are theoretically less interesting. We aim to counter this bias by also investigating the alternation at low levels of abstraction, which means we will require a large corpus to guaranty a sufficient amount of instances of the alternating variants.

Investigating an alternation at a low levels of abstraction also comes with an important methodological advantage, however, as argued by Grondelaers, Speelman and Geeraerts (2008: 158–160). In practice, it means that the linguistic context of the variants is restricted to a high degree, e.g. by keeping the verb constant across all instances. As a result, the instances under scrutiny become highly comparable, and essentially form series of "minimal pairs" (Grondelaers, Speelman and Geeraerts 2008: 158). This allows a crystalline view of what motivations remain to determine the choice between both variants.

The goal of answering the where-question is mainly to show that it is possible to investigate argument structure alternation at several levels of abstraction in a systematic way and illustrate one way how it may be done.[8] To summarize, the how-, why- and where-questions are related as follows: the why-question will yield hypotheses to be tested for the how-question, and the how-question functions as an empirical criterion for the where-question.

## 1.4    Structure of the thesis

This thesis is structured as follows. **Chapter 2** lays the theoretical groundwork for the study. To do so, it first discusses the several definitions of an alternation that are employed in various subdisciplines of linguistics, as well as their corresponding

---

material scraped from the internet. At the moment of reading, it has probably grown notably larger than 7.9 billion.

[8] Argument structure alternations are called argument alternations in the remainder of this thesis, for reasons of brevity.

perspectives on grammar, and positions this study within the field of alternation studies. Next, it turns to Dutch grammar description to present the evolution and definition of the prepositional object. Finally, it provides an overview of the literature on the Dutch transitive-prepositional alternation.

**Chapter 3** delineates the alternation and describes the extraction of the data from the corpus. It begins by presenting the employed corpus, viz. the Sonar-corpus of written Dutch, and details the alternating variants in terms of the Alpino-parses used in the corpus (van Noord 2006; Oostdijk et al. 2013a). It describes the selection of the alternating verbs, the extraction of the alternating instances, and it sketches the various levels of abstraction at which the alternation may be determined.

**Chapter 4** is devoted to the why-question. It formulates the a priori hypotheses of this study, classified into the three groups presented in Section 1.2. First, it discusses why and how we expect lectal distinctions to be relevant for the alternation. Here, we will concentrate on the distinction between the Netherlandic and Belgian regiolects of Dutch (Grondelaers, Speelman and Geeraerts 2008; Grondelaers and van Hout 2011).[9] Next, the chapter turns to semantic hypotheses, where the focus lies on the lexical origin hypothesis (Goldberg 1999, 2006; Perek and Lemmens 2010). This hypothesis describes a mechanism of how argument constructions – or constructions in general – can acquire meaning. Finally, we hypothesize what influence language processing may have on the alternation. Here, we zoom in on the mechanisms that have been proposed to underlie the Complexity Principle of Rohdenburg (Rohdenburg 1996; Ferreira and Dell 2000; Fenk and Fenk-Oczlon 1993; Hawkins 2004).

**Chapter 5** details the manual selection of the alternating instances. First, it sets out the demarcation lines between categorical contexts, alternation factors and knock-on effects. Next, its lists all instances that were removed for the entire dataset, as well as those that were retained after some consideration, and then does the same for the individual alternating verbs.

**Chapter 6** tests the predictions based on the lectal and semantic hypotheses formulated in Chapter 4. Since these predictions pertain to several levels of abstraction, it starts with those at the highest level and then works its way down. The behavior of the verbs *verlangen* 'desire', *peilen* 'gauge' and *zoeken* 'search' is explored in depth.

**Chapter 7** extends the search for alternation factors between the variants by employing a hypothesis-generating or data-driven procedure. The focal point here lies still with lectal and semantic differences between the variants, since we want to be able to properly control for these in Chapter 8. The procedure is based on Memory-Based Learning (Daelemans and van den Bosch 2005), which will be used to yield hypotheses that are in turn tested through manual annotation.

---

[9] The avoid confusion, we will use the adjective *Netherlandic* to refer to the country of the Netherlands, and *Dutch* to refer to the Dutch language.

**Chapter 8** tests the hypotheses based on the influence of language processing. It first checks whether the rationale set out in Chapter 4 still holds for the present dataset, and then applies the predictions to the data and interprets the results.

**Chapter 9** finally provides answers to the questions how, why and where the Dutch transitive and prepositional-intransitive argument structures vary. It reiterates some of the issues discussed in Chapter 2 and applies them to the alternation under scrutiny, and sketches a number of avenues for future study. It concludes this thesis by summarizing its contributions to various fields of study.

# 2    Theoretical background and previous research

This chapter begins by theoretically grounding two concepts that are key to the investigation: the alternation and the prepositional object. Section 2.1 provides an overview of the various outlooks on alternations that are at hand in the literature, and elucidates to which the present study subscribes. Section 2.2 is concerned with the historical development and the synchronic definition of the prepositional object. Next, Section 2.3 surveys the available literature on alternations between the transitive and prepositional intransitive construction in Dutch and some closely-related languages. There is no specific section devoted to research on formally similar alternations, like the Dutch transitive-reflexive alternation (Pijpops and Speelman 2017) or the English, Dutch and German dative alternations (Bresnan et al. 2007; Colleman 2009a; De Vaere, De Cuypere and Willems 2018), although such work is of course discussed in passing throughout this thesis. For overviews of work on the dative alternations in English, Dutch and German, see respectively Röthlisberger (2018a: 11–24), Geleyn (2017: 68–72) and Adler (2011: 179–190).

## 2.1    What is an alternation?

In the literature on linguistic alternations, various perspectives can be found on what exactly constitutes an alternation. This section contains a succinct survey of these perspectives and their corresponding notions of grammar. The main goals of this section are (i) to prevent terminological confusion by trying to delineate each perspective as clearly as possible from the others; and (ii) to discuss some of the concerns that have been raised with the corpus-based study of alternations. Still, we do not mean that these are the only possible perspectives on alternations, nor that they are incompatible with one another. In fact, at the end of this section, we make explicit that the present study is compatible with two perspectives on

alternations, and also explain how the study can still make a useful contribution when any of the other perspectives are taken.

We start with three definitions that are complementary to one another: the classic sociolinguistic alternation, the classic psycholinguistic alternation, and the classic 'grammatical' alternation. In principle, the first two definitions could be collapsed as 'two forms with the exact same meaning', but since they stem from two distinct research traditions and are employed quite differently, we present them seperately.

The classic sociolinguistic alternation is defined in Labov (1972a,b), as consisting of two or more formal realizations that vary across the (socio)lectal dimension, without a referential meaning difference between the variants (Lavandera 1978; Robinson 2012: 38–39). Typical examples would be cases of phonetic variation, such as the pronunciation of the suffix -*ing* as [ɪn] or [ɪŋ] or the realization of the *r* in *fourth* (Labov 1966a,b).

Meanwhile, the classic psycholinguistic alternation consists of two forms that do not vary according to the lectal or semantic dimension, but according to processing-related factors (e.g. Elsness 1984; Ferreira and Dell 2000). Typical examples would be language structures where one or multiple words are optional, such as *The coach knew (that) you missed practice* or *The astronauts (who were) selected for the mission* (Ferreira and Dell 2000: 297; see the overview in Ferreira and Schotter 2013: 1548–1551).

Finally, the 'grammatical' alternation concerns two constructions that exhibit a systematic difference in form, that typically corresponds to a systematic difference in semantics, be it in terms of referential meaning, aspect, construal or otherwise. Lectal and processing-related factors are either not at play or not considered of primary concern to the researcher (Levin 1993; Hanks 2013: 186–207; Broekhuis, Corver and Vos 2013: 401–594). The term 'grammatical alternation' is taken over from Arppe et al. (2010: 12). They use the term to refer to alternations whose study originated in formal syntax, although the study of this type of alternation have since long been exported to other fields such as cognitive linguistics, partially in order to argue against the claims of formal syntax (Arppe et al. 2010: 12–13). Typical examples would be the English locative alternation, as (10), the English conative alternation as in *Paula hit (at) the fence* (Broccias 2001; Hanks 2013: 205) or the Dutch causative-state alternation as in (11) (van Hout 1996: 51). For more examples of grammatical alternations in English, see Levin (1993: 25–110).

10. a. *The farmer loaded apples into the cart.*
    b. *The farmer loaded the cart with apples.*
                    (taken over from Levin 1993: 2)

11.  a.   *De foto     hangt aan de muur.*
          the  picture  hangs on   the wall
          'The picture hangs on the wall.'

     b.   *Joep hangt de foto     aan de muur.*
          Joep hangs the picture  on    the walls
          'Joep hangs the picture on the wall.'
                              (taken over from van Hout 1996: 62)

These three definitions are consistent with a modular approach to language, where the language system or 'core grammar' can readily be separated from language use and language variation, and where each of these can and even should be studied separately in their respective subdisciplines of syntax, psycholinguistics and sociolinguistics (Geeraerts 2010a: 73–77). However, the three definitions are more difficult to uphold in a 'recontextualized' approach to language, where the language system is as much a product of language use as vice versa, and social meaning, such as prestige or signals of group membership, are considered just as 'core' to grammar as semantic meaning (Geeraerts 2010a: 77–95). This is the approach taken by cognitive, usage-based approaches (Geeraerts 2006a; Bybee 2010). The argument would be that an accurate description of language would need to take into account both the social and semantic meaning of expressions, and to study this meaning requires linguists to investigate language use (Geeraerts 2006a: 2–6; Grondelaers, Geeraerts and Speelman 2007: 149–152). Hence, the language system, language use and language variation should not be investigated separately, but should rather be studied in an integrated fashion (Geeraerts 2006b: 30–31, 2010a).

In principle, a first possibility would be to simply relax the three classic definitions by allowing some influence of the other types of factors, as long as this influence is minor. For instance, two formal variants would then still be said to constitute a sociolinguistic or psycholinguistic alternation even if there is a meaning difference between both, as long as this meaning difference is small and not overly obvious in each instance of the variants.

A second option, one that takes the concerns of cognitive and usage-based linguistics more seriously, is to redefine alternations as choice points of the individual language user (Bresnan 2007; Bresnan et al. 2007). This choice should not be regarded as a conscious choice, but more generally as the "process of selection from any set of alternative outcomes arising in a process." (Wallis forthc.: 2–3). The focus on the individual would correspond to a notion of grammar as the "cognitive organization of one's experience with language" (Bybee 2006: 711), or the "internal representation of existing regularities" (Geeraerts, Kristiansen and Peirsman 2010: 5) or "I-grammar" (Zuidema and Verhagen 2010: 54).

While this grammar itself is often considered inherently probabilistic in nature (see Röthlisberger 2018a: 3 and references cited there), its probabilities may in practice be so biased that no real choice is left (Bresnan 2007; Bresnan and Hay 2008; Röthlisberger 2018a: 3-5, 53-58). That is, if any one factor or a combination of factors would be too dominant in determining the appearance of the formal variants, this would effectively rob the individual from his/her choice. In such a case, the variants could not be said to constitute a probabilistic alternation.

Once we adopt this definition and take a potential alternation under scrutiny, we need to answer two questions: who is the individual and what exactly constitutes his/her choice? This first question is not necessarily trivial. Geeraerts (2010b: 66–68) and Verhagen (2013) argue that the concept of a representative language user is no longer tenable in usage-based and cognitive linguistics, and, indeed, empirical research has unveiled outspoken differences in alternations between language users. What is a clear probabilistic choice for one group of language users may be more of a categorical distinction for another group (Grondelaers, Speelman and Geeraerts 2008; Speelman 2014: 519). Moreover, even within a single community of language users, there may be outspoken differences between the language users (Grondelaers and Speelman 2008; Dąbrowska 2015a; Fonteyn and Nini 2018).[10]

The second question, viz. what exactly constitutes his/her choice, requires us to delineate the alternation. It can be split into two further sub-questions: (i) how many formal variants are in competition; and (ii) when exactly does a factor become too dominant, or where does probabilistic become categorical? The first subquestion is the primary concern of the Principle of Accountability (Labov 1969: fn. 20; Tagliamonte 2012: 19–21; Van de Velde 2013: 164–165; Szmrecsanyi et al. 2016).

For example, Pijpops and Van de Velde (2018b) attempt to find out which factors determine the alternation of the Dutch partitive genitive construction which has an optional -s ending, as in (12). Of course, the choice of the language user in (12) is not limited to the forms with or without -s ending. Other options include *een interessant iets* 'an interesting something', *iets dat interessant is* 'something that is interesting', *iets dat me interesseert* 'something that interests me', *iets waarvoor ik me interesseer* 'something that interests me', etc.[11] Hence, any occurrence of *iets dat interessant is* etc. presents a choice point where the language user could have opted for *iets interessants* or *iets interessant*, yet Pijpops and Van de Velde (2018b) do not include these occurrences in their dataset.

---

[10] Also see Petré and Van de Velde (2018) on individual differences in language change.

[11] Incidentally, these last two options would instantiate the psych verb alternation introduced in (7)-(8) (Pijpops and Speelman 2017).

12. *Ik zoek   geen tv-werk, maarals iets       interessant(s)     zich aanbiedt*
    I   search no   tv-work, but   if   somethinginteresting(-gen) itself presents
    'I'm not looking for TV work, but if something interesting presents itself,...'
                          (taken over from Pijpops and Van de Velde 2018: 104)

Furthermore, there is the issue of synonymous partitive genitives that are actually taken up in the dataset as additional instances, such as *wat interessants, wat interessant, iets boeiend, wat boeiends,* etc (all: 'something interesting'). Surely, the language user in (12) could have opted for these forms as well. Should they therefore be regarded as additional variants, rather than simply as other instances of the partitive genitive variants? This "Pandora's box" of additional variants is discussed by Szmrecsanyi et al. (2016: 25–26) regarding the English genitive alternation, and they urge for more discussion on the topic.[12]

The second sub-question, viz. when exactly does a factor become too dominant, or where does probabilistic become categorical, has mostly been discussed with reference to semantic factors and dubbed the problem of semantic equivalence (Geeraerts, Kristiansen and Peirsman 2010: 7–9, Grondelaers and Speelman 2007; this problem has deep roots in Labovian sociolinguistics, starting with at least Sankoff 1973, Lavandera 1978, Romaine 1984, Silva-Corvalán 1986 and Jenny Cheshire 1987). Still, the question equally applies to lectal and processing-related factors. As for lectal factors, consider the variants *mij-mich.* Both are object pronouns used to refer to the first person, and at first sight, it seems immediately obvious what determines the choice between both: when speaking Dutch, one should use *mij*, when speaking German *mich*. Clearly, this distinction is so categorical that the forms cannot qualify as an alternation.

Or can they? There is no strict binary distinction between the Continental West-Germanic dialects, and the state border separating the Dutch and German standard languages is historically contingent (van der Wal and van Bree 2008). In fact, *mich* functions as the dialectal variant in large areas of Belgium and the Netherlands, notably in the Limburgian provinces, where *mij* is the standard (Kruijsen and van der Sijs 2016). In these areas, the *mij-mich* choice could be

---

[12] The methodological difficulties of investigating an alternation with more than two variants are substantial (Arppe et al. 2010: 13), but this does not or should not bear on the discussion whether or not it is theoretically desirable to do so. Several strategies of dealing with multiple variants have already been proposed. These include the use of multinominal regression (Arppe 2008; Gries 2013: 322–324; Levshina 2015: 277–290) or the clever combination of multiple binomial regression models (Szmrecsanyi et al. 2016; Wallis forthc.: 4).

probabilistically determined by social and situational factors.[13] The question is then in which contexts these factors become so stringent that no real choice is left for the language user.

For instance, the author of this thesis grew up in Brabant, just west of the transitional zone between Brabantic *mij* and Limburgian *mich*, about 60 km from the state border separating the Dutch and German standard languages. He would, albeit rarely to scarcely ever, use *mich* when talking to his aunt and mother. When talking to colleagues in a formal setting however, *mich* would be completely out of the question. Still, having recently moved to the Belgian province of Limburg, the probability of *mich* in his language use is poised to rise – even if only from 'hardly ever' to 'seldom'. Would his occurrences count as genuine instances of the *mij-mich* alternation, and if so, in which situations and during which periods of his life?[14]

A parallel reasoning can be made for processing-related factors. Consider the influence of verb frequency in the competition between the Germanic strong and weak inflection, e.g. Dutch *dolf* vs. *delfde* 'delved'.[15] While many verbs exhibit both variants, the strong variant tends to dominate among the highly frequent verbs (Lieberman et al. 2007; Carroll, Svare and Salmons 2012; De Smet and Van de Velde 2019). For instance, the Dutch verb *zijn/wezen* 'be' is so frequent that it is

[13] Other features that originated in these areas, more specifically in the ethnically mixed *Cité*-neighborhoods, have in fact acquired social meaning and spread across the entirety of Dutch-speaking Belgium (Marzo 2016; Marzo, Zenner and Van de Mieroop 2016).

[14] This example is not as exceptional as it may seem at first sight. Consider the variants *arms-weapons* (Partridge 1966; Gramley and Pätzold 2004: 29). Before 1066, these had a categorical geographical distribution: *arms* being used South of the Channel, and *weapons* North. After the Norman conquest of England, this became a categorical sociolectal distribution: *arms* being used by Norman nobility, *weapons* by Anglo-Saxon peasants. Gradually, this developed into the semi-categorical semantic distinction that is in place in contemporary English. Still, while this development was in progress, the choice would have been probabilistically driven by both sociolectal and semantic factors. The same development has occurred for large parts of the current English lexicon (Gramley and Pätzold 2004: 29).

On an individual level, Petré et al. (forthc.) e.g. show that Margaret Cavendish only employed the English alternation between the third person verbal endings *-s* and *-th* during a specific period of her life, while categorically using *-s* in other periods. It is hard or even impossible, however, to a priori demarcate the exact period when she did use the alternation probabilistically.

[15] The weak-strong distinction transects the regular-irregular distinction, with many weak verbs having become irregular (e.g. English *buy-bought*) , and many strong verbs still being regular, especially in Dutch and German (e.g. English *sing-sang, drink-drank*) (De Smet and Van de Velde 2019). An additional complicating factor in this alternation is that in the West-Germanic languages, the simple past tense as such is also coming into competition with the present perfect (Trost 1980; Abraham and Conradie 2001; Drinka 2004; Vanmassenhove, Du and Way 2017).

unquestionably conjugated strongly as *was* 'was' rather than weakly as *\*weesde* 'was', just like its English cognate. Clearly, this verb could not be said to alternate. But where exactly should we draw the line between verbs categorically not exhibiting the weak variant, and those verbs probabilistically unlikely to do so? This question becomes especially pressing since verbs that used to be conjugated strongly in a categorical fashion, or were at least thought to do so, have in fact been observed to occasionally exhibit the weak inflection and vice versa (van Santen 1997; Knooihuizen and Strik 2014; De Smet and Van de Velde 2019). It could be argued that at least theoretically, any verb has the potential to be weakly conjugated, and hence presents a choice point in the sense of Wallis (forthc.: 2-3, and see above). Of course, the decision of how strictly to demarcate the list of alternating verbs will have a direct impact on the results, as excluding too many verbs of high frequency from the analysis would lead to an underestimation of the influence of verb frequency on the alternation.

Another example would be contexts in which the language user is under stringent pressure to express his/her message in as few words as possible, as in the context of telegrams, tweets or newspaper headlines. The Daily Mirror of May 2, 1945 headlined *U-Boat chief claims he's new Fuehrer*. Given the limited space available to the journalist, would he/she have been able to write *U-Boat chief claims that he's new Fuehrer*? Considering that almost all function words are absent from this headline, perhaps including *that* was simply not an option.[16]

Or consider discourse formulas such as *I think, I guess* etc. as in (13). Some researchers argue that these instances have grammaticalized as epistemic markers to such a degree that *that* is no longer possible, and as a consequence, they should be kept out of the analyses (e.g. Thompson and Mulac 1991, also see Torres Cacoullos and Walker 2009 and the overview in Shank, Bogaert and Plevoets 2016a: 32–44). Still, other contend that the probability of *that* in these instances is so low precesely because a following subordinate clause is highly predictable, and hence they form outstanding examples to substantiate the influence of predictability, processing load or early mention on *that*-omission (e.g. Ferreira and Dell 2000; Roland, Elman and Ferreira 2006: 270; Jaeger 2010).

13.   *I think exercise is really beneficial, for anybody.*
          (taken over from Thompson and Mulac 1991: 313)

---

[16] It could be argued that this example is a matter of register or genre rather than processing. See Subsection 4.3.1.2 on why we include channel restrictions under the processing-related variables.

Turning to semantic factors, the usual cutoff-point between the probabilistic and the categorical is whether both variants share the same truth conditions (e.g. Bresnan and Ford 2010: 170–172; Röthlisberger 2018a: 55; see the overview in Röthlisberger 2018a: 17–23).[17] This is a bit peculiar from a theoretical point of view, given the emphasis of cognitive linguistics on prototype categories with fuzzy boundaries and its dismissal of strictly referential models of language (Geeraerts 2010a: 81; also see Lakoff 1987: xi–xvii, 157-228; Geeraerts 2010c; Lemmens 2016: 91–92). Still, truth conditions could function as a practical demarcation line in the slippery field of semantics (Glynn 2010: 240; Röthlisberger 2018a: 18–19).

Even in practice though, it is not always clear which instances do and do not share the same truth conditions, as lexical senses are rarely if ever discrete categories (Geeraerts 1993; Kilgarriff 1997; Glynn 2010). For an example, we return to the Dutch psych verb alternation, which occurs not just with the psych verbs *ergeren* 'annoy' and *interesseren* 'interest', as an (7)-(8), but also with *storen* 'disturb', as in (14). One of the hypotheses of Pijpops and Speelman (2017), the so-called Agentivity Hypothesis, predicts that more agentive stimuli promote the use of the transitive construction.

Now, Pijpops and Speelman (2017) exclude a number of instances of transitive *storen* from their analysis, stating that these refer to physical rather than psychological actions, and hence categorically fail to exhibit the reflexive construction. These instances include a number of clearly physical meanings, but also instances like (15)-(16), where they identify the lexical sense 'to interrupt someone, who is working or talking to someone else, by asking for their attention'. It could be argued, however, that this sense of *storen* does not strictly refer to a different action than in (17), i.e. that they are not strictly referentially different from (17). Instead, the reason why instances like (15)-(16) so strongly prefer the transitive variant might just be because their stimulus is outspokenly agentive. In that case, excluding these instances would lead to an underestimation of the influence of semantic factors on the alternation, in favor of e.g. processing-related ones.[18]

14.  a.  *Al te chauvinistisch gedrag     stoort   mij.*
         All too chauvinistic      behavior   disturbs  me
         'Behavior that is too chauvinistic disturbs me.'
                        (Transitive construction, stimulus-subject)

---

[17] The use of truth conditions as cutoff-point is an inheritance from sociolinguistics (Labov 1972a: 271).

[18] Pijpops and Speelman indeed find that processing factors outperform semantic factors in their regression model, but also warn not to base any conclusions on this finding regarding the relative importance of semantics and processing in the alternation (Pijpops and Speelman 2017: 239).

b.　*Ik stoor me aan al te chauvinistisch gedrag.*
I　disturb me to　all too　chauvinistic　behavior.
'Behavior that is too chauvinistic disturbs me.'
>　(Reflexive construction, experiencer-subject; taken over
>　from Pijpops and Speelman 2017: 230)

15.　*Goedenmiddag, met softwarehouse Been, sorry dat ik u stoor.*
good-afternoon with software-house Been　sorry that I　you disturb
'Good afternoon, this is the software company Been, sorry to bother you.'
>　(Transitive construction, taken over from Pijpops and
>　Speelman 2017: 225)

16.　*Stoorde ik je weer tijdens de uitzending gisteren hahahaha*
Disturb I　you　again　during the broadcast　yesterday hahahaha
'Did I disturb you again during the broadvast yesterday hahaha'
>　(Transitive construction, taken from the ConDiv corpus,
>　Grondelaers et al. 2000, file-id: #HOLL_2.SML)

17.　*Stoor je niet aan mij. Ik luister wel toe.*
Disturb you not to　me I　listen PART PART
'Don't let me disturb you, I'm just listening.'
>　(Reflexive construction, taken from the ConDiv corpus,
>　Grondelaers et al. 2000, file-id: fn000559.pos)

Note that this is essentially the same problem as we encountered above for the *mij-mich* alternation, for the highly frequent verbs in the Germanic weak-strong alternation and for discourse formulas in the English *that*-alternation.[19] The problem can be stated more generally as follows: being strict in the demarcation of an alternation for one type of variable while being more lenient towards other types could lead to an over- or underestimation of the influence of one type of variable in linguistic variation. Robinson (2012) and Van de Velde, Franco and Geeraerts (forthc.) argue that semantics might get the short end of the stick in this regard, precisely because of its slippery nature. This would be a most unfortunate consequence, given that cognitive linguistics "sees language [...] as something primarily semantic" and takes an explicit interest in meaning (Geeraerts 2006a: 3).

These questions, i.e. who is the individual and what exactly constitutes his/her choice, may be answered with lesser or greater difficulty depending on the

---

[19] For more examples of this problem in the English and Dutch dative alternations, see Goldberg (1995: 141–179), Gries and Stefanowitsch (2004), Colleman (2010), Lehmann and Schneider (2012), Bernolet and Colleman (2016).

alternation at issue. For instance, concerning the English genitive alternation, it is clear that it presents a probabilistic choice for all – or at least most – English-speaking individuals, and it seems relatively easy to demarcate the non-alternating occurrences of the *of*- and *s*-genitives (cf. the overview in Heller 2018: 54–57). Still, the issue of additional formal variants does present a potential hiccup (Szmrecsanyi et al. 2016), as does the status of semantic factors (Stefanowitsch 2003). For the English dative alternation, the questions already become harder to answer than for the genitive alternation (cf. Röthlisberger 2018b, and the overview in Röthlisberger 2018a: 53–58), just as for the alternation under scrutiny in this thesis (see Chapter 5 and Subsection 9.3.1).

Still, most researchers agree that whatever the alternation, these questions are merely practical issues that can be overcome when given proper caution (Tagliamonte 2012: 3–22; Wallis forthc.). For instance, D'Arcy (2014: 225, 227) proposes that if one of the levels of a categorical or ordinal independent variable exhibits a proportion of 95% or more for one of the variants, then that level should be considered a categorical context, and its instances should be excluded from the analysis.[20]

By contrast, Arppe et al. (2010) argue that the questions present more serious theoretical concerns. To fundamentally deal with these concerns, they propose to broaden the definition of alternations from choice points of the individual language user to practical research setups created by the researcher in order to test hypotheses (Arppe et al. 2010: 12–15). Concretely, a researcher would start from a hypothesis, choose two or more forms about which the hypothesis makes different predictions, and then study these forms as an alternation to test the predictions.

This final definition essentially corresponds to a generalization of the classic psycholinguistic definition in two ways. First, the hypothesis or hypotheses in question may also pertain to lectal and semantic differences, instead of only to processing-related ones. Second, the choice between the forms is not necessarily made by the individual language user.

For instance, the verb *to milk* is invariably conjugated weakly in English, as *milked*, rather than strongly as e.g. *\*malk* (cf. *sing ~ sang, drink ~ drank*). As such, the individual English language user does not have a choice between the weak and strong inflection for this verb. Still, the community of English language users did have a choice at one point, as it is of course entirely possible to use *malk*. That is, the community could have just as well chosen strong *malk* as their norm rather than weak *milked*. In fact, the communities of Dutch and German language users did actually choose to conjugate the cognate verb *melken* strongly as *molk* 'milked' – although weak *melkte* 'milked' is also in use (de Vriendt 1965: 245; Haeseryn et

---

[20] D'Arcy (2014) refers to Guy (1988: 130), but Guy only discusses distributions between independent variables, or *factor groups*, not between an independent and the dependent variable.

al. 1997: 91; Duden 2009: 489). The alternation between strong and weak inflection for the Germanic cognates of the verb *milk* can then still be studied to test a hypothesis that e.g. the strong inflection would more rapidly recede in areas with and during periods of increased language contact (e.g. De Smet et al. 2017; Pijpops, Beuls and de Velde 2018).

A similar example can be found with the English alternation between synthetic and analytic expression of the comparative and superlative, e.g. *subtler, subtlest* vs. *more subtle, most subtle* (see D'Arcy 2014 and references cited therein). The community of English language users seems to have already decided for most adjectives which variant should be used, and this choice has even been ossified in prescriptive works such as the OED (2014), leaving hardly any choice for the individual language user. Still, it could be studied what alternation factors determine this choice made by the community.

When applied to corpus research, this hypothesis-driven perspective fits in with a population-level perspective on grammar where the individual no longer takes central stage. Instead, grammar would be viewed as the "observable regularity in the language use realized by a specific community" (Geeraerts, Kristiansen and Peirsman 2010: 5; Geeraerts 2010d; Steels 2000, 2011a; Beckner et al. 2009; Dąbrowska 2015b; Van de Velde 2017) or a description of "E-language" (Zuidema and Verhagen 2010: 54).

This alternative definition takes away the need to answer the first question, viz. who is the individual, as the individual is no longer crucial to the definition of an alternation nor of grammar. As for the second question, which concerned the selection of the data, this would be determined by the interest of the researcher, and would ideally be dictated by either the hypotheses at issue or the theory underlying them or, less ideally, by practical concerns in the research design.

Note that this does not imply a more permissive approach to data selection. In fact, this would make it easier to justify the exclusion of instances such as (15)-(16) as part of a conservative research design to answer the Agentivity Hypothesis. The rationale would be that, in order for us to confirm the Agentivity Hypothesis, we demand it to still yield a significant effect even among those instances where differences in agentivity are not overtly clear.[21] In addition, the drive for a conservative research design could also be used to justify the exclusion of instances based on the researcher's personal intuitions, as is done in e.g. Szmrecsanyi et al. (2016: 4–5) and Pijpops and Van de Velde (2016: 565–566). As a result, it would not require the researchers to identify their own intuitions as representative for the

---

[21] The distinction between (15)-(16) and (17) could also be reasoned to be an aspectual one, rather than one in terms of agentivity. Again, the exclusion of (15)-(16) could then be justified by a desire to conservatively test the Agentivity Hypothesis. By contrast, under a view of alternations as individual choices, it would not be immediately clear why a researcher should allow semantic differences in terms of agentivity into his/her dataset, while weeding out all aspectual differences.

individual language user, which is highly problematic in a cognitive, usage-based framework (Geeraerts and Kristiansen 2015: 370, Wolk et al. 2013: 386).[22] In other words, the view on alternations as researcher's setups could result in a data selection procedure that is in practice identical to that of the view on alternations as an individual's choice points.

Alternation studies in variational and cognitive linguistics are generally compatible both with the view of alternations as researcher's setups, and with the view of alternations as choice points of individuals. While many studies only deal with lectal and processing-related factors (Bresnan and Hay 2008), others do focus on semantic factors (see e.g. Speelman and Geeraerts 2009; Speelman 2014: 519–529; Levshina and Heylen 2014; Perek 2015 and several contributions in Glynn and Fischer 2010; Glynn and Robinson 2014). The present study will test both processing, lectal, and semantic hypotheses, and the selection of the data will be determined by these hypotheses, following a conservative research design. This study is hence also compatible with the last two views on alternations. Still, since the corpus components that will be employed in the present study contain no information on individual language users, we will strive to err on the side of caution in that regard, and limit our conclusions regarding individuals to what we deem necessary to explain the tendencies that we find in language usage (cf. Subsection 9.3.2).

Should the reader disagree with these definitions of an alternation, and prefer a classic or relaxed sociolinguistic, psycholinguistic or grammatical definition, then the present study can still be relevant. In that case, we would ask to read this study as an attempt to determine for which verbs and objects the transitive and prepositional variants constitute an actual alternation and for which they do not, as is done in e.g. Condamines (2018). In Subsection 9.3.1, we list the segments of the alternation where it exhibits a (semi-)systematic difference in meaning, those segments for which there is an important lectal distinction, and those segments which can be used in research on language processing.

Adherents of the (classic) grammatical definition can take the first group and shove aside the others as not 'really' being part of the language system, but merely instantiating uninteresting language variation and usage. Adherent of the classic sociolinguistic definition can focus on the second group, and those of the psycholinguistic definition would be most strongly interested in the third group. Still, we hope to illustrate that studying all three dimensions – semantic, lectal, and processing – in an integrated fashion does lead to a more accurate and complete understanding of the functioning of language (Grondelaers, Geeraerts and Speelman 2007; Arppe 2009; Geeraerts 2010a).

---

[22] Researcher's intuitions can still be useful and even crucial in drafting hypotheses, of course (Grondelaers, Geeraerts and Speelman 2007: 150–151; Geeraerts 2010b).

There are two more points of debate concerning alternations that deserve to be mentioned here, which cross-cut the definitions. The first is whether the factors determining an alternation are represented in the minds of the individual language users, even at times when they are not actively producing language in a natural setting (Theijssen 2012: 123–132; Divjak, Dąbrowska and Arppe 2016; Klavan and Divjak 2016). This is only necessary under the grammatical definition of alternations. As for a sociolinguistic alternation, it would be possible that a member of one community of language users is simply unaware of the existence of another linguistic variant in another community, or that he/she is aware of the existence of the other variant, but remains ignorant about its social meaning.[23] In that case, information regarding the distribution between both variants would not be stored in the mind of that particular language user.

Under a psycholinguistic definition, it would be in principle similarly possible that, while producing language, processing limitations or channel restrictions directly cause language users to e.g. more often produce *that* with certain matrix verbs, even if such lexical preferences are not registered in the mind (Jaeger 2010; Ferreira and Schotter 2013). This last example would also hold for the definition of alternations as choice points of the individual (cf. above and Wallis forthc.: 2–3 on the term 'choice'), and both examples would fall under the definition of alternations as a researcher's setups. We will return to this point in Subsection 9.3.2 for the alternation factors under scrutiny in the present study.

A second point of debate is whether and which alternations are entitled to a separate theoretical status, i.e. whether the variants of an alternation have a specific theoretical relation to one another that is qualitatively different from their relation to other constructions. In generative approaches to language, this separate theoretical status would typically be that of transformations, while in cognitive linguistics, the term *allostructions* has been proposed by Cappelle (2006) for the English particle placement alternation. In the meantime, the same case has been made for the English dative and locative alternations by Perek (2012, 2015: 145–174). Again, this is an issue that can only be determined from one alternation to the next. We will hence return to this in Subsection 9.3.3, where we discuss whether the transitive and prepositional variants may be called allostructions.

---

[23] For example, a Netherlandic language user of Dutch once admitted in personal communication to the author that he was not aware of the existence of the variant without *-s* ending in the Dutch partitive genitive construction, e.g. *iets leuk(s)* 'something fun', studied in Pijpops and Van de Velde (2016), Pijpops and Van de Velde (2018b). This variant is widely used in Belgium, and even occasionally occurs in the Netherlands as well.

## 2.2     What is a prepositional object?

The grammatical status and precise delineation of the prepositional object has led to a number of vivid discussions in the tradition of Dutch grammar description. To understand why, it is crucial to understand the object's historical development. This section hence starts by giving a short overview of this development, and then turns to the discussions regarding the definition of the prepositional object.

### 2.2.1.     Historical development of the prepositional object

This subsection is chiefly based on van der Horst (2008), and van der Horst and Van de Velde (2008). The development of the prepositional object is part of the story of the evolution of the preposition itself. The first prepositions are thought to have originated from locative adverbs and perhaps even earlier from nouns (Dal 1962: 49; Heine and Kuteva 2002: 47, 2007: 83; Ringe 2006: 64–65). For instance, in sentences such as (18), *ad* is assumed to have originally functioned as an autonomous locative adverb with an intensifying meaning – and hence could be omitted in early Latin – when it was reanalyzed as having a grammatical relation to the following noun and forming a single, prepositional constituent with it. The first prepositional constituents were then adverbial adjuncts. This is how we still find them in the earliest records of Old Dutch, as in (19).

18.  *Copias  (ad)  urbem ducere*
     Troops  (to)  city    lead
     'To lead troops to the city.'
                    (taken from van der Horst and Van de Velde 2008: 57)

19.  *… ande so  an      themo anaginna  zueyuet  sie*
          and  so  from   the     beginning destroy  them
      '… and so destroy them from the beginning.'
                    (The Leiden Willeram, taken over from Sanders 1971: 45.5,
                    cited in van der Horst 2008: 244)

In Middle Dutch, two new developments took root. The earliest was that the preposition could now be used to link two nouns to each other, i.e. prepositional constituents began to be used as attributive adjuncts, as in (20). The latter development is of greater interest to the present study, however: the preposition

also began to be used to link a verb with one of its complements, as in *ane hem* 'on him' in (21). That is, we begin to see the first prepositional objects. Among the earliest verbs to have appeared with prepositional objects are *geloven (aan)* 'believe' and *haken (naar)* 'strive'. Crucial in this step, and what first set the prepositional object apart from the prepositional adjunct, was that the preposition was used to introduce complements of the verbs that were hitherto exclusively realized by bare nominal constituents.[24] Typically, the bare nominal realization and the realization as prepositional object would co-exist for an extended time. Still, the link with the prepositional adjunct remained clear in that the choice of preposition was often still variable, at least to a some degree.

20.  *Vant  hi  thoeft      van  eenen  doden man*
     found  he  the_head  of    a          dead  man
     'He found the head of a dead man.'
                         (Gerritsen and Wilmink 1994: line 138; cited in van der
                         Horst 2008: 534)

21.  *Nochtan so  gheloefden  vele    liede uten       ghemeinen  volke*
     still      so  believed      many  folks from_the  ordinary      people
     *ane  hem.*
     on    him
     'Still, many of the ordinary people believed him.'
                         (de Bruin 1970: line 144; cited in van der Horst 2008: 471)

The big break-through of the prepositional object coincides with the dissolution of the case system in spoken Dutch around the 16th and 17th centuries. We see an outspoken increase in the number of verbs appearing with a prepositional object, and more verbs seem to settle on a choice of preposition, which in turn leads to a semantic bleaching of that preposition. This increase is continued in later centuries.

   During the development of the prepositional object, its relation to the verb seems to be growing ever stronger. In fact, it even appears to be overtaking its relation to the nominal complement in prominence. When several verbs with a prepositional object are coordinated, the preposition may stay with the verb rather than with the nominal complement, as in (22). The nominal complement may remain entirely unexpressed while the preposition is retained, as in (23), or the preposition may be stranded in front of the second verbal pole, as in (24).[25] Similar

---

[24] Bare nominal, but with case marking of course.

[25] See Subsection 5.1 for a short introduction to the Dutch sentence structure, which revolves around two verbal poles. It should be noted, however, that sentences such as (24) are rare,

developments are simultaneously taking place for prepositional attributive adjuncts regarding its relation to the first versus the second noun.


22.  *Over  de aanpassing van de procedure is (…)  overleg     gevoerd*
     about  the adaptation  of    the procedure is  (…)  consultation done
     *met en   instemming verworven van  de betrokken bieders.*
     with and  permission obtained    of    the involved  bidders
     'The involved bidders have been consulted about the adaptation of the
     procedure, and their permission was obtained.'
                     (NRC Handelsblad 14 september 2002, cited in van der
                     Horst and Van de Velde 2008: 60)


23.  *Het streven naar is  belangrijker     dan het bezitten     van.*
     the  striving to   is  more_important  than the possessing  of.'
     'Striving for something is more important than possession it.'
                     (taken over from van der Horst and Van de Velde 2008: 58–
                     59)


24.  *Bananen houd ik niet van.*
     Bananas  love I  not  of
     'I don't like bananas.'
                     (taken over from van der Horst 2008: 462)


The maturing of the prepositional object does not mark the end of the evolution of the preposition. For one, new prepositions keep emerging in Dutch, both from adverbs, such as *aldus* 'thusly', and from nouns, such as *richting* 'direction' (van der Horst 2012). Again, these prepositions first occur with adverbial constituents, but it is likely that prepositional objects will follow. An indication that such a development is already incipient can be found in the sentence in (25), found on the internet.


25.  *Informeer richting   uw   mogelijkheden.*[26]
     inform    direction  your possibilities
     'Inform yourself as to your possibilities.'

and many native speakers of Dutch would call them incorrect language. We did not encounter any in our dataset.

[26] Taken from https://radioalkmaar.nl/2019/02/22/voorkom-een-burn-out/ (accessed 3 July 2019). *Informeren* is listed as a verb that normally takes a prepositional object with *naar* in Haeseryn et al. (1997: 1172), Smedts and Van Belle (2003: 284) and Vandeweghe (2004: 102).

## 2.2.2.　　Definition of the prepositional object

To define the prepositional object, a number of criteria are listed in reference grammars and journal articles that primarily aim to delineate it from its formally most similar neighbors, the prepositional adjunct and the prepositional indirect object (a.o. van den Toorn 1971; Klooster, Verkuyl and Luif 1974: 41−44; Haeseryn et al. 1997; Balk-Smit Duyzentkunst 2000: 185−191; Smedts and Van Belle 2003: 282−284; Vandeweghe 2004: 95−104; Broekhuis 2004: 102−109; de Swart 2014: 451−455). Most of these criteria, listed below, are based on the preposition having a strong connection to the verb. The example sentences are taken over from Schermer-Vermeer (1988: 12), Vandeweghe (2004: 96−99), and Broekhuis (2004: 107−108).

- The preposition cannot be replaced by another preposition as in (26), or only to a limited degree for some specific verbs, as in (27).
- The preposition cannot be intensified, as in (28).
- The preposition cannot be omitted, as in (29).
- The preposition cannot be stressed, as in (30).
- The preposition does not have its own locative or temporal meaning, as in (31).
- The preposition cannot be separated from the verb, as in (32)-(33), while it can be separated from the nominal complement, as in (34) and (24).
- The prepositional object is normally placed behind prepositional adjuncts, i.e. closer to the second verbal pole, as in (35).

Further criteria include:

- The participant expressed by the prepositional object cannot be referred to by an adverb, unless that adverb contains the preposition, i.e. is a prepositional adverb, as in (36).
- The prepositional object may be a subordinate clause, as in (37).

26.　a.　*Zij　legde　zich　neer　bij　/　\*onder die　beslissing.*
　　　　she　layed　herself down　with / \*under that　decision
　　　　'She accepted that decision.'
　　　　　　　　(Prepositional object)

　　b.　*Zij legde　zich　neer　bij　/　onder die　appelboom.*
　　　　she layed　herself down　with / under that　apple_tree
　　　　'She layed down next to / underneath that apple tree.'
　　　　　　　　(Prepositional adjunct)

27.  *Denk aan / over / om de boterhammen.*
     Think on / about / for the sandwiches
     'Don't forget the sandwiches.'
                    (Prepositional object)

28.  a.  *Hij stond op / \*boven op zijn rechten.*
         he stood on / above on his rights
         'He insists on his rights.'
                    (Prepositional object)

     b.  *Hij stond op / boven op een zeepkist.*
         he stood on / above on a soapbox
         'He stood on / on top of a soapbox.'
                    (Prepositional adjunct)

29.  a.  *Zij heeft zich aan haar man / \*haar man geërgerd.*
         she has herself on her husband / \*her husband annoyed
         'Her husband annoyed her.'
                    (Prepositional object)

     b.  *Zij heeft aan haar man / haar man een cadeautje gegeven*
         she has on her husband / her husband a present_DIM given
         'She has given to her husband / her husband a little present.'
                    (Indirect object)

30.  a.  *\*Hij stond óp een eerlijke behandeling.*
         he stood ón a fair treatment
         'He insisted ón a fair treatment.'
                    (Prepositional object)

     b.  *Hij stond óp de stoel.*
         he stood ón the chair
         'He stood on the chair.'
                    (Prepositional adjunct)

31.  a.  *Jan vecht voor een betere wereld.*
         Jan fights for a better world
         'Jan fights for a better world.'
                    (Prepositional object)

b. *Jan   vecht   voor  de  school.*
Jan   fights   for   the school
'Jan is fighting in front of the school.'
(Prepositional adjunct)

32. a. *\*Wachten mag  niet  op  zijn  vader.*
Waiting   may not   on  his   father
'Waiting is not allowed on his father.'
(Prepositional object)

b. *Wachten  mag  niet  op  het perron.*
waiting    may not   on  the platform
'It is not allowed to be waiting on the platform.'
(Prepositional adjunct)

33. a. *\*Marie wacht en   zij doet dat  op haar vriend.*
Marie waits  and  she does that  on her  friend
'Marie is waiting and she does that on her friend.'
(Prepositional object)

b. *Marie  wacht en   zij doet dat  op het station.*
Marie  waits  and  she does that  on the station
'Marie is waiting, and she is doing that in the station.'
(Prepositional adjunct)

34. a. *Op wie   Jan  wacht, is  zijn  vader.*
on whom Jan  waits  is  his   dad
'For whom Jan is waiting, is his dad.'
(Prepositional object)

b. *\*Waarop  Jan   wacht, is  het perron.*
whereon  Jan  waits  is  the platform
'On which Jan waits, is the platform.'
(Prepositional adjunct)

35. a. *…  dat  Jan  op het perron  op  zijn  vader  wacht.*
that  Jan  on the platform on his   dad    waits
'… that Jan is waiting for his dad on the platform.'
(Adjunct precedes object)

b.  *... dat  Jan  op zijn  vader  op het perron  wacht.*
that Jan  on his   father  on the platform waits
'... that Jan is waiting for his dad on the platform.'
(Objects precedes adjunct)

36.  a.  *Natasja  wachtte op  zijn  komst.*
Natasja waited   on his   arrival
'Natasja awaited his arrival.'
(Prepositional object)

b.  *Natasja  wachtte erop.*
Natasja  waited   thereon
'Natasja awaited it.'
(Prepositional object)

c.  *Natasja  wachtte op  het perron.*
Natasja  waited   on  the platform
'Natasja waited on the platform.'
(Prepositional adjunct)

d.  *Natasja  wachtte er.*
Natasja  waited   there
'Natasja waited there.'
(Prepositional adjunct)

37.  a.  *Zij  legde zich  erbij     neer  dat die  beslissing genomen*
she  layed  herself therewith down  that that decision    taken
*werd.*
was
'She accepted that that decision was taken.'
(Prepositional object)

b.  *\*Zij legde zich  eronder    neer  dat die  appelboom...???*
she  layed  herself thereunder  down  that that  apple_tree
'She layed down that that apple tree...'
(Prepositional adjunct)

Objections can be raised against all of these criteria, however (Schermer-Vermeer 1988, 1990, 1991, 2006, 2007). For one, the existence of the alternation under scrutiny in this thesis is already in violation with the third criterion, viz. that the preposition cannot be omitted. For another, Footnote 26 already mentioned that the

preposition *naar* may be replaced with *richting* when introducing the prepositional object of the verb *informeren* 'inform', which runs counter to the first criterion.

In other words, the definition of the prepositional object is not watertight, as also explicitly recognized by some scholars (den Hertog 1903: 65–66; Duinhoven 1989: 54). This has led to a longstanding discussion regarding the status of the prepositional object, with some researchers arguing that prepositional objects should not be recognized as a separate syntactic category at all (for an overview, see Greve 2009: 44–66). Instead, they should be classified among the prepositional adjuncts – albeit possibly as a specific type of prepositional adjunct, in view of their strong relation to the verb (Kollewijn 1899; Mars 1969; Perridon 1991; Schermer-Vermeer 1988, 1990, 1991, 2006, 2007).

We agree with the critics of the prepositional object that there is a continuum between the prepositional adjunct and the prepositional object, and that there are many instances where the criteria that have been proposed to recognize prepositional objects give conflicting results or engender disagreement between language users. Yet, admitting that there is a grey zone between two categories is not the same as abandoning the distinction between the two altogether. The general consensus still holds that the prepositional object constitutes a separate syntactic category, and we follow this consensus (o.a. Zwaan 1972; van den Toorn 1971; Duinhoven 1989; Broekhuis 2004; Broekhuis 2007).

Still, there is no consensus among researchers on where exactly to draw the line between prepositional adjunct and prepositional object. For one example, Duinhoven (1989: 49) actually considers *naar je paraplu* 'to your umbrella', from our very first example sentence (1b), which is repeated below, to be a prepositional adjunct, while Haeseryn et al. (1997: 1168) and Balk-Smit Duyzentkunst (2000: 186) present it as a clear example of a prepositional object.

For another example, the status of prepositional constituents such as (38) is controversial. Vandeweghe and colleagues argue that these present a subtype of the prepositional object, viz. partner objects, that are distinct from more prototypical prepositional objects, which they call content objects, as in (39) (Vandeweghe and Devos 2003; Vandeweghe 2004: 99–100; Vandeweghe 2011a; Vandeweghe 2011b; Colleman and Delorge 2010; Vandeweghe and Colleman 2011; Colleman 2014). This division is to mirror the division among nominal objects between indirect and direct objects, respectively. Conversely, Broekhuis and colleagues view prepositional constituents as in (38) as mere prepositional adjuncts, and only consider the so-called content objects to be actual prepositional objects (Broekhuis 2004; Broekhuis, Corver and Vos 2013: 284–329; Broekhuis 2014).

1.   b.   *Zoek je    naar je    paraplu?*
          search you   to    your umbrella
          'Are you looking for your umbrella?'
                         (taken over from Haeseryn et al. 1997: 1168)


38.   *Jan  praat  met  Els.*
      Jan  talks  with Els
      'Jan is talking to Els'.
                         (taken over from Vandeweghe 2011: 95)


39.   *Zij  praten over   X.*
      they talk    about  X
      'They are talking about X.'
                         (taken over from Vandeweghe and Devos 2003: 106)


Duinhoven (1989: 41) argues that the fundamental reason for these continued disagreements is that the category of the prepositional object has only recently emerged and is still very much in flux, as shown in the previous subsection, and particular constituents are simply caught somewhere in the transition between the prepositional adjunct and the prepositional object, at least in the present stage of the Dutch language. As such, one will always be forced to make some essentially arbitrary choices in delineating the prepositional object.

Still, we will of course require such a delineation when gathering the data. The developers of the Alpino-parser adhere to the consensus of reference grammars in regarding the prepositional object as a separate syntactic category, and use the criteria listed above when drawing a line between the prepositional object and other prepositional constituents. As far as we are aware, the development of the Alpino-parser constitutes the only exhaustive attempt at distinguishing Dutch prepositional objects from adjuncts. We will hence take over this delineation for reasons of consistency and practicality, but explicitly accept and expect it to be arbitrary to some degree (see Section 3.2 and Footnote 92 in Subsection 9.3.1).


## 2.3     Alternations between transitive and prepositional intransitive variants


As mentioned above, there has not yet been an in-depth corpus investigation of the Dutch transitive-prepositional alternation. Still, the alternation is mentioned in passing by a number of studies. The goal of these works is not to find the factors

governing the alternation, but rather, for example, to account for the L1-acquisition of the lexicon-syntax interface (van Hout 1996), to examine the status of the prepositional object (Broekhuis 2004), to discuss Differential Object Marking in Dutch (de Swart 2014), etc.[27] These studies only talk about subtle meaning differences between the variants. Distinctions in terms of language processing or lectal dimensions are not mentioned, as far as we are aware. In other words, the alternation has only been considered under the definition of a grammatical alternation (cf. Subsection 2.1).

We will use the term *agent* to refer to the participant performing the action expressed by the verb, and *theme* to refer to the participant with which the action expressed by the verb is concerned, which may be realized in a nominal or prepositional constituent, depending on the choice of variant. These terms are only meant as  practical designators, however, and we do not mean to attribute any specific theoretical status to them, either as a semantic roles or similar concepts.

Van Hout (1996: 50–53, 94–98, 118–120) proposes that the prepositional variant puts the focus on the action expressed by the verb, while the transitive variant puts the focus on both the action and the result (cf. Dixon 1991: 280). What this contrast in practice amounts to, however, would depend on the verb (van Hout 1996: 120). For instance, for some verbs, such as *bouwen (aan)* 'build' and *eten (van)* 'eat', it would involve a distinction between a telic interpretation for the transitive variant and a atelic interpretation for the prepositional variant. For others, such as *duwen (tegen),* the question would be whether the theme participant is affected or not.

The notion of affectedness is interpreted very broadly in van Hout (1996: 120). For one example, it would also include conation or attempted action, even though a successful action does not entail that the theme argument is actually affected (cf. de Swart 2014: 460). This is proposed for a.o. *schoppen (naar)* 'kick' (van Hout 1996: 50-51, 119-120). For another, affectedness would also hold for *bekijken* vs. *kijken naar* 'look at', where transitive *bekijken* allegedly involves a more precise and serious look at the theme argument than prepositional *kijken naar* (van Hout 1996: 120). Other verbs, such as *vertrouwen (op),* are also mentioned, but it is not specified what exact meaning differences would be instantiated for them.

Van Voorst (1996: 235–236, 241–242) mentions a number of additional meaning distinctions. For *kauwen (op)* 'chew', as in (40), the transitive variant would imply that the gum or the meat is being chewed for the purpose of feeding oneself. Meanwhile, for the verbs *bijten (op)* 'bite' and *trappen (tegen)* 'kick', the

---

[27] Van Hout (1996: 97) and Broekhuis, Corver and Vos (2013: 562) call the alternation the *transitive-oblique alternation*, while Broekhuis (2004: 121) calls it *NP-PP-alternations*, and de Swart (2014: 455) calls it (a paradigmatic type of) *DOM alternation*, where DOM stands for Differential Object Marking. We prefer the name *transitive-prepositional alternation*, because the meaning of the term *oblique* is theory-dependent and the term *prepositional* is more transparent, while the terms *NP-PP alternations* and *DOM alternations* could be interpreted more broadly than the alternation under scrutiny here.

use of the transitive variant with animate themes would imply an act of communication or punishment, while the use of the transitive variant with the theme *een bal* ' a ball', as in (41), would mean that the ball is being made part of a play. These implications are said not to arise for their prepositional counterparts. The distinction for *trappen (tegen)* 'kick' is explained in terms of affectedness: by warning or punishing someone, or by making a ball part of a play, these themes would be impacted internally.

40.  a.  *Zijn vlees kauwen*
         his   meat chew
         'To chew one's meat.'

     b.  *Op zijn vlees  kauwen*
         on his   meat   chew
         'To chew on one's meat.'

     c.  *Zijn kauwgum     kauwen*
         his   chewing_gum chew
         'To chew one's chewing gum.'

     d.  *Op zijn kauwgum     kauwen*
         on his   chewing_gum chew
         'To chew on one's chewing gum.'
                     (taken over from van Voorst 1996: 236)

41.  a.  *Zij   trapte de inbreker.*
         she   kicked the burglar
         'She kicked the burglar.'

     b.  *Hij trapte tegen    de inbreker.*
         he  kicked against   the burglar
         'He kicked against the burglar.'

     c.  *Zij   trapte de bal.*
         she   kicked the ball
         'She kicked the ball.'

     d.  *Hij trapte tegen    de bal.*
         he  kicked against   the ball
         'He kicked against the ball.'
                     (taken over from van Voorst 1996: 241)

Broekhuis (2004: 122) also states that "the meaning shift can go in all sorts of ways" (our translation). Affectedness is also mentioned, but for verbs such as *eten (van)* 'eat', the question would be whether the action expressed by the verb involves the theme argument as a whole or only a part of it. For *geloven (in)* 'believe in' in (42), the use of the transitive variant would mean that Jan believes what Marie is saying, while the prepositional variant would mean that Jan has faith in Marie. For *verlangen (naar)* 'desire', the meaning difference would be something along the lines of 'demand' for the transitive variant and 'long for' for the prepositional variant (also see Van de Velde 2014a: 340). The discussion of the alternation in Broekhuis, Corver and Vos (2013: 562–564) is based on van Hout (1996) and likewise refers to telicity, affectedness and conation.

42.  a.  *Jan  gelooft  Marie.*
Jan   believes Marie
'Jan believes Marie.'

 b.  *Jan  gelooft  in  Marie.*
Jan   believes in  Marie
'Jan believes in Marie.'
(taken over from Broekhuis 2004: 122)

De Swart (2014: 456–463) only deals with the alternation among verbs that express contact, such as *krabben (aan)* 'scratch' or *knijpen (in)* 'pinch'. He asserts that the relevant distinction is that the theme is sentient in the transitive variant, and non-sentient in the prepositional variant. That is, that the theme is capable of feeling the scratching or pinching in the transitive variant but not in the prepositional variant. Still, he also concedes that this distinction doesn't hold for all contact verbs. For *snijden (in),* he refers to affectedness, while for others, such as *aaien (over)* 'carress' and *likken (aan)* 'lick' as in (43), the transitive variant is allegedly limited to sensual contexts when the theme is inanimate.

43.  a.  *Als   hij haar het geld  heeft overhandigd, likt ze haar duim*
when  he  her  the money has  handed_over  licks she her  thumb
*en   wijsvinger   en  telt   het na.*
and index_finger  and  counts it  PART
'When he has handed the money over to her, she licks her thumb and index finger and counted it.'
(WR-P-P-B-0000000160.p.2344.s.4)

   b.    *Ze   likte   aan  haar vinger en    streek*  een  wenkbrauw glad
         she  licked on    her   finger and brushed an    eyebrow      smooth
         'She licked her finger and smoothened an eyebrow.'
                        (WR-P-P-B-0000000007.p.223.s.1)

Since these studies have other goals than discovering the factors determining the transitive-prepositional alternation, the proposed meaning distinctions between the variants are merely asserted, not tested. The questions also remain how these meaning distinctions would have developed, and how the same preposition can apparently induce different meaning shifts for different verbs, e.g. *naar* for *schoppen* 'kick' and *verlangen* 'desire'. To sum up, the current picture of transitive-prepositional alternations in Dutch is somewhat underspecified and rather diverse, with notions like affectedness being called upon to account for various meaning distinctions, alongside differences in terms of completeness, sentience, sensuality and additional verb-specific distinctions.

    Similarly, when investigating a German prepositionsal intransitive construction called the *search*-construction, Proost (2015, 2017) is compelled to define this construction a priori in semantic terms in order to distinguish it from a number of formally identical argument constructions that are posited to be different in meaning. Even then, the alleged meaning of the construction, viz. 'prospective possession', is said to give rise to different concrete semantic effects for various verbs. In fact, for some verbs such as *suchen* 'search', the meaning of 'prospective possession' is noted not to generate any clear meaning distinction to the transitive variant at all (Proost 2017: 24, 33–45).

    A comparable picture can also be found in Lenci (2012), who attempts a systematic investigation of Italian argument alternations, as in (44)-(45).

44.  a.   *L'    assemblea ha deciso   l'  acquisto   della    società.*
          The  assembly  has decided  the purchase  of_the  company
          'The assembly decided the purchase of the company.'

     b.   *L'    assemblea ha deciso   sull'  acquisto   della   società.*
          The  assembly  has decided  on_the purchase  of_the company
          'The assembly decided on the purchase of the company.'
                        (taken over from Lenci 2012: 14)

45.  a.   *Gianni  ha rimproverato suo  padre  per   questo.*
          John     has reproached  his  father for    this
          'John reproached his father for this.'

b.  *Gianni ha   rimproverato questo a   suo padre*
    John  has  reproached   this   to his  father
    'John reproached his father for this.'
              (taken over from Lenci 2012: 14)

Lenci notes the following.

> *As we said, something like "direct involvement" seems to be at stake here, and this might suggest that "affectedness" is again the crucial semantic factor. However, this interpretation would require us to stretch the meaning of affectedness well beyond its standard (fairly high) vagueness and polysemy, thereby impairing its reliability as a truly explanatory notion in semantics.* (Lenci 2012: 14)

Perek (2014) formulates essentially the same critique regarding a proposal by Dixon (1991: 280) about the English conative alternation, as in (46). Dixon states that the transitive variant puts the focus on the effect of the action, while the prepositional variant puts the focus on the subject's engagement in the action (cf. above, van Hout 1996: 52).

46.  a.  *Sam chipped at the rock.*        (taken over from Broccias 2001: 77)
     b.  *Sam chipped the rock.*

Perek remarks the following.

> *While this account seems reasonable at first blush, such an abstract characterization must still go a long way towards the actual semantic contribution with individual verbs, leaving a heavy burden to processes of meaning construction.* (Perek 2014: 64)

The same point is made more generally by Dąbrowska (2017: 21–38). She argues that the use of vague – and hence elusive and moldable – semantic notions prevents researchers from formulating strict predictions, and thus hampers the application of hypothesis-testing in linguistics. This is not only a practical problem when doing research, however, but also a theoretical one. To illustrate this, we turn to the English conative alternation, which is probably the best-studied alternation between a transitive and prepositional intransitive construction (o.a. Levin 1993: 6–10, 41–42; van der Leek 1996; Broccias 2001; Perek and Lemmens 2010; Perek 2014).

Under compositionality, the meaning of an utterance should be equal to the sum of the meaning of its constituting elements (Goldberg 1995: 13–16). For example,

we can form the utterance in (46a) by combining the lexical *chip*-construction with the abstract conative construction, as well as a number of other constructions that are not at issue here, such as the morphological paste tense construction and the lexical *rock*-construction. When doing so, we should be able to calculate or predict the meaning of the entire utterance by summing up the meanings of these constructions. Assuming that the meaning of the lexical *chip*-construction is 'to break off small fragments' (Oxford English Dictionary, lemma *chip*) and the meaning of the conative construction is 'direct action at' (Pinker 1989: 104; Goldberg 1995: 63; Perek and Lemmens 2010), we get the compositional meaning 'Sam directed action at and broke off small fragments from the rock'.

However, the meaning of the utterance in (46a) is more specific than that. It also includes that the action of chipping is repeated or takes place in a bit-by-bit fashion (Broccias 2001: 77; Perek 2014: 76–77). This bit-by-bit meaning chunk has to be contributed by the conative construction, because it is no longer present in the utterance when we replace the conative construction by the transitive construction, but retain all other constructions, as in (46b).

The sensible solution then seems to update the meaning of the conative construction, and make it more specific, e.g. to replace its meaning 'direct action at' by 'repeatedly act on'. However, this would be fallacious, since this meaning is not present when the conative construction is combined with the verb *strike*, as in (47). Here, the conative construction leaves unspecified whether physical contact was made (Perek 2015: 134).

47.  *He asked, 'What's the problem?' and in the same moment, struck at the man stiff-armed, a karate blow.*
     (British National Corpus, corpus-id: FS8-1809, cited in Perek 2015: 134)

The knowledge that the man is not necessarily hit in (47), while the chipping was repeated in (46a) needs to be stored somewhere. If it cannot be stored for the abstract conative construction, where can it then be stored? This corresponds to the where-question presented in Section 1.3. Perek (2014, 2015: 105–142) deals with this problem by investigating the alternation at a lower levels of abstraction, viz. by distinguishing between several verb-class specific constructions. Still, while verb-class-specific constructions are already more concrete than fully abstract argument constructions, they still rank rather high on the abstraction scale. What if we want to investigate the alternation at even more concrete levels, e.g. where the verb-slot is fixed? We return to this question in Section 4.2. In the next chapter, we turn to the extraction of the data.

# 3    Data extraction

This chapter presents the extraction of the data used in the corpus analysis. Section 3.1 contains a description of the employed corpus, viz. the Sonar-corpus of written Dutch, and explains why this corpus was chosen. Section 3.2 then describes how the alternating verbs were selected, and Section 3.3 describes the extraction of the instances. Finally, Section 3.4 sketches the various levels of abstraction at which this alternation can be investigated, ranging from fully abstract argument constructions to concrete instances.

## 3.1    The corpus

The Sonar-corpus may be viewed as the written counterpart to the Corpus of Spoken Dutch. The Corpus of Spoken Dutch contains 10 million words of spoken language from Belgium and the Netherlands, distributed over several components that correspond to various registers, and aims to be representative for spoken standard Dutch (Oostdijk et al. 2002). Similarly, the Sonar-corpus contains 500 million words of written Dutch from Belgium and the Netherlands and is divided over several components, including subtitles, books, tweets, chat material and newspapers (Oostdijk et al. 2013a). Its aim is to form a representative cross-cut of written standard Dutch.

The Sonar-corpus was chosen for three reasons. The first and most crucial reason is its large size. A large corpus is required not so much because we necessarily want large datasets, but because we want specialized datasets (cf. Grondelaers, Speelman and Geeraerts 2008: 158). For example, we want to be able to fix the verb to *peil* 'gauge' and the theme to *reactie* 'reaction' and still find sufficient data to do some sort of analysis. The second reason is that the corpus has been syntactically annotated with the Alpino-parser (van Noord 2006). These parses will prove crucial when selecting the alternating verbs in Section 3.2, when preparing the dataset for manual checking in Section 3.3, and when formulating the hypotheses in Chapter 4. The third reason is that the corpus aims to be representative for written Standard Dutch. As such, we can claim that our findings will likely be representative for written Standard Dutch as well.

The Sonar-corpus does come with an important disadvantage, though. The distribution of Belgian and Netherlandic material over the components is uneven,

as can be seen in Table 1, which presents an overview of the size of the Belgian and Netherlandic material in each of the Sonar-components. Table 1 also shows that the majority of the data comes from fairly formal registers, most notably from newspapers and periodicals and magazines. This was unavoidable though, since it was decided to prioritize the quality of the syntactic parses over the use of more varied material, and the quality of Alpino parses is notably better for more formal registers. In fact, the material of the four most informal Sonar-components was purposely excluded from this study, since the quality of their Alpino-parses was a priori deemed too low (Oostdijk et al. 2013b: 49–50). These components are the text messages, chats, tweets and discussion lists. This is also the reason why the total size in the lower right corner of Table 1 does not quite reach 500 million.

| Component code | Component name | Belgium | The Netherlands | Belgium or the Netherlands (unknown) | Total |
|---|---|---|---|---|---|
| WR-P-E-C | e-magazines | 6,336,962 | 2,289,286 | 0 | 8,626,248 |
| WR-P-E-E | electronic newsletters | 0 | 0 | 1,917 | 1,917 |
| WR-P-E-F | press releases | 28,246 | 304,549 | 0 | 332,795 |
| WR-P-E-G | subtitles | 18,680,798 | 0 | 9,529,048 | 28,209,846 |
| WR-P-E-H | teletext pages | 448,865 | 0 | 0 | 448,865 |
| WR-P-E-I | web sites | 2,018,109 | 1,079,231 | 14,249 | 3,111,589 |
| WR-P-E-J | wikipedia | 0 | 0 | 23,001,184 | 23,001,184 |
| WR-P-E-K | blogs | 139,765 | 0 | 0 | 139,765 |
| WR-P-P-B | books | 48,581 | 26,134,187 | 2,013 | 26,184,781 |
| WR-P-P-C | brochures | 1,065,231 | 44,959 | 103,192 | 1,213,382 |
| WR-P-P-D | printed newsletters | 0 | 33,529 | 0 | 33,529 |
| WR-P-P-E | guides manuals | 16,386 | 212,579 | 7,134 | 236,099 |
| WR-P-P-F | legal texts | 21,458 | 292,862 | 10,375,361 | 10,689,681 |
| WR-P-P-G | newspapers | 152,288,524 | 59,381,224 | 0 | 211,669,748 |
| WR-P-P-H | periodicals magazines | 79,495,036 | 13,563,888 | 0 | 93,058,924 |
| WR-P-P-I | policy documents | 107,223 | 32,285 | 8,572,043 | 8,711,551 |
| WR-P-P-J | proceedings | 238,285 | 75,740 | 0 | 314,025 |
| WR-P-P-K | reports | 561,077 | 1,631,642 | 25,504 | 2,218,223 |
| WR-U-E-E | written assignments | 0 | 357,947 | 0 | 357,947 |
| WS-U-E-A | auto cues | 25,268,159 | 2,819,822 | 0 | 28,087,981 |
| WS-U-T-B | texts for the visually impaired | 0 | 675,082 | 0 | 675,082 |
| Total | | 286,762,705 | 108,928,812 | 51,631,645 | 447,323,162 |

Table 1: Number of words from Belgium and the Netherlands in each the Sonar-components employed in the present study.

Furthermore, only the data for which the country of origin is known will be used in the analyses, because we will want to distinguish between Belgian and Netherlandic occurrences. This means that the components of the electronic newsletters and Wikipedia will not be used. Only when building distributional vectors will all of the data in Table 1 be employed (see Chapter 4). Further information of the Sonar-corpus and its Alpino-parses can be found in Oostdijk et al. (2013b), van Noord et al. (2013) and van Noord, Schuurman and Bouma (2011).

## 3.2    Selection of the verbs

In the nomenclature of the Alpino XML-parses, an occurrence of the transitive construction is defined as a parent node with at least two child-nodes: one child-node with the POS-tag *verb* and the REL-tag *hd*, indicating that it is the syntactic head of the parent node, and another with the REL-tag *obj1*.[28] No other child-node with an object-relation may be present. By object-relation, we mean a REL-tag with the value *obj1, ld, me, obcomp, obj2, pc, predc, se*, or *vc*. Likewise, an occurrence of the prepositional intransitive construction is defined as a parent node with at least one child-node with the CAT-tag *verb* and the REL-tag *hd*, and another child with the REL-tag *pc* or a child with the REL-tag *ld* and the CAT-tag *pp*. This means that we consider as prepositional objects both those constituents marked as a *prepositional complement*, and the prepositional constituents marked as a *locative or directional complement*. Again, no other child-node with an object-relation may be present. The presence of a subject was not required, such that imperatives would also be included.[29]

We then needed a list of verbs that may alternate between both constructions, i.e. that can express the same participant as either a direct object or as a prepositional object. To obtain such a list, we employed a procedure similar to Lenci (2012: 4−9). First, all verbs were selected that appeared both in the transitive construction and the prepositional intransitive construction. To narrow this list

---

[28] The values of the Alpino-tags mentioned in this paragraph have the following meanings: *hd*: syntactic head, *obj1*: direct object if in a verb phrase, or nominal constituent within a prepositional phrase, *ld*: locative or directional complement, *me:* measure complement, *obcomp*: comparative complement, *obj2*: indirect object, *pc*: prepositional complement, *predc*: predicative complement, *se*: obligatory reflexive object, *vc*: verbal complement, *pp:* prepositional constituent.

[29] When two or more verbs were coordinated with the same object, these were viewed as multiple occurrences of a transitive or prepositional intransitive construction. When multiple objects were coordinated with the same verb, this was viewed as one occurrence of the transitive or prepositional intransitive construction.

down, we looked at the root of the deepest syntactic head of the theme, hereafter referred to as *theme root* (cf. Lenci 2012: 5–7, also see Zeldes 2013: 270).[30] It was then required that at least three theme roots appeared both in the transitive and the prepositional intransitive variant for each unique combination of a verb and a preposition. This is a fairly admissive criterion (cf. Lenci 2012: 5–7), but it is still possible that infrequent alternating verbs have not passed the bar. As such, the obtained list is certainly not intended to be exhaustive: it is only meant as a starting point of investigation. This selection left us with 778 verbs, or 2360 unique combinations of verb and preposition.

Next, an expert survey was composed for 200 randomly selected verbs, which yielded 650 unique combinations of a verb and a preposition. For each of these combinations, four annotators independently judged whether the combination presented a genuine alternating verb, or whether it should be discarded as a false positive. These annotators all held university-level degrees in Dutch linguistics and live in central northern Belgium. To help with these decisions, two general criteria and six example sentences were given. The first criterion was semantic in nature: it stated that the direct and prepositional object needed to realize the same participant, and should be able to do so for an in principle infinite number of theme roots. In other words, both variants needed to be interchangeable and productive. This was to exclude combinations such as *zingen (over)* in (48), where both variants are not interchangeable, or *luisteren (naar)* in (49), as the transitive variant of this verb is limited to a number of themes, such as *radio* 'radio' and *muziek* 'music'. The second criterion was formal in nature and stated that there should not be any additional complements, thereby excluding e.g. instances of the dative alternation. As example sentences, three pairs of a transitive and a prepositional example sentence were selected such that each pair had the same theme root.

48.　a.　*Ik zing een liedje.*
　　　　I　sing　a　little_sing
　　　　'I'm singing a song.'

　　b.　*Ik zing over een liedje.*
　　　　I　sing　about　a　little_song
　　　　'I'm singing about a little song.'

---

[30] In Alpino terms: we recursively selected the child-node with the REL-tag *hd*, resolving indexation and conjunction. For the transitive occurrences, the direct object constituent itself was used as starting point; for the prepositional intransitive construction, the nominal child-node of the prepositional object, viz. the child-node with REL-tag *obj1*, was used. Finally, we looked at the ROOT-tag of the deepest head. If the ROOT-tag was not available, one of the values of other tags with a similar function was used. Instances with complement clauses were ignored, for now.

49.　a.　　*Ik　luister　radio.*
　　　　　 I 　 listen 　 radio
　　　　　 'I'm listening to the radio.'

　　　b.　　*Ik　luister　naar　de　radio.*
　　　　　 I 　 listen 　 to 　　 the radio
　　　　　 'I'm listening to the radio.'

The agreement between the annotators was substantial with Cohen's κ = 0.705 (Landis and Koch 1977).[31] This was judged to be sufficiently high for one annotator, viz. the author, to judge the remaining combinations. In this way, 101 verbs or 121 unique combinations were finally selected. These are the combinations in Table 2.

| *Aan* 'on' | *Naar* 'to' | *Tegen* 'against' |
|---|---|---|
| **Constructive group** | **Telephonic group** | **Collision group** |
|   *bouwen* 'build' |   *bellen* 'ring' |   *drukken* 'press' |
|   *knutselen* 'tinker' |   *opbellen* 'ring up' |   *duwen* 'push' |
|   *schaven* 'plane' |   *telefoneren* 'phone' |   *rammen* 'ram' |
|   *timmeren* 'carpenter' | |   *schoppen* 'kick' |
| | **Motoric group** |   *stampen* 'stomp' |
| **Monastic group** |   *graaien* 'grasp' |   *tikken* 'tick' |
|   *gehoorzamen* 'obey' |   *grabbelen* 'scramble' |   *trappen* 'kick' |
|   *geloven* 'believe' |   *grijpen* 'grab' | |
|   *vasthouden* 'hold on' |   *happen* 'snap' | **Other** |
|   *verzaken* 'forsake' |   *schoppen* 'kick' |   *spreken* 'speak' |
|   *weerstaan* 'resist' | | |
| | **Venatic group** | *Tot* 'to' |
| **Tractional group** |   *jagen* 'hunt' |   *spreken* 'speak' |
|   *krabben* 'scratch' |   *vissen* 'fish' | |
|   *likken* 'lick' | | *Uit* 'from' |
|   *rammelen* 'rattle' | **Other** |   *citeren* 'cite' |
|   *tillen* 'lift' |   *peilen* 'gauge' | |
|   *trekken* 'pull' |   *verlangen* 'desire' | *Van* 'of' |
| |   *zoeken* 'search' |   *drinken* 'drink' |
| ***Other*** | |   *eten* 'eat' |
|   *ruiken* 'smell' | *Om* 'around' |   *proeven* 'taste' |
|   *ontspringen* 'spring from' |   *dribbelen* 'dribble' |   *snoepen* 'eat sweets' |
|   *verhelpen* 'remedy' | |   *vreten* 'devour' |
| | *Onder* 'under' | |
| |   *dienen* 'serve' | |
| *Bij* 'near' | | |
|   *dienen* 'serve' | | |

---

[31] This may be considered reasonably high, since κ-values tend to underestimate agreement when the distribution between the labels is skewed (Viera and Garrett 2005: 262–263). This is the case here: of the 2360 combinations to be judged, only 122 were finally accepted.

| *Door* 'by' |
|---|
| *roer* 'stir' |

**In** 'in'

**Intrusive group**
*infiltreren* 'infiltrate'
*knijpen* 'pinch'
*rammen* 'ram'
*roeren* 'stir'
*snijden* 'cut'
*woelen* 'tumble'

**Other**
*dealen* 'deal'

**Langs** 'along'
*strijken* 'brush'
*vegen* 'sweep'
*wrijven* 'rub'

**Met** 'with'

**Displacement group**
*gooien* 'throw'
*knoeien* 'bungle'
*manoeuvreren* 'manoeuvre'
*morsen* 'spill'
*schuiven* 'slide'
*sjouwen* 'lug'
*slepen* 'drag'
*slingeren* 'sling'
*smijten* 'fling'
*strooien* 'scatter'
*werpen* 'hurl'
*wuiven* 'wave'
*zeulen* 'haul'
*zwaaien* 'swing'
*zwiepen* 'swish'

**Reciprocal group**
*huwen* 'wed'
*knuffelen* 'hug'
*neuken* 'fuck'
*telefoneren* 'phone'
*trouwen* 'marry'

**Op** 'on'

**Taction group**
*bijten* 'bite'
*bonken* 'bang'
*drukken* 'press'
*duwen* 'push'
*kauwen* 'chew'
*klikken* 'click'
*krabben* 'scratch'
*krassen* 'scrape'
*rammen* 'ram'
*schoppen* 'kick'
*slaan* 'hit'
*terugslaan* 'hit back'
*tikken* 'tick'
*wrijven* 'rub'

**Venatic group**
*jagen* 'hunt'
*vissen* 'fish'

**Other**
*bezuinigen* 'economize'
*gelijken* 'resemble'
*vertrouwen* 'trust'

**Over** 'over'

**Caressing group**
*aaien* 'caress'
*krabben* 'scratch'
*strijken* 'brush'
*vegen* 'sweep'

**Other**
*berichten* 'message'
*regeren* 'rule'

**Voor** 'for'
*boeten* 'pay'
*kiezen* 'choose'
*vrezen* 'fear'

**Compositional verbs**
*inboeten (aan)* 'lose'
*meedoen (aan)* 'participate'
*meedoen (met)* 'participate'
*binnendringen (in)* 'penetrate'
*voorgaan (in)* 'conduct'
*opklimmen (tegen)* 'climb'
*aanrijden (op)* 'collide'
*aanrijden (tegen)* 'collide'
*afrollen (van)* 'roll down'
*meespelen (met)* 'play along'
*binnentreden (in)* 'enter'
*binnenvallen (in)* 'burst into'
*aanvatten (met)* 'start'

Table 2: Verbs selected as alternating between the transitive and prepositional intransitive construction.

When the combinations are ordered according to their preposition, it becomes easy to discern a number of semantic groups of verbs, as is done in Table 2. Under the compositional verbs, those verbs are grouped that incorporate an erstwhile preposition as a particle (e.g. *binnendringen*, lit. 'inside_press'). Apparently, this incorporation has progressed to such a degree that a new preposition often crops up (e.g. *binnendringen in*, lit. 'inside_press in'). Under *other*, all verbs are listed which cannot readily be assigned to a certain group.

## 3.3 Extraction of the instances

All instances of the verbs in Table 2 were extracted from the corpus. This was done by selecting each node with the corresponding ROOT-tag, e.g. *boet_in* for the verb *inboeten* 'lose', and the POS-tag *verb*. Each instance of a verb is hence counted as a separate observation. This means that when multiple objects were coordinated with the same verb, this was viewed as a single occurrence of the transitive or prepositional intransitive construction. Each verb was extracted together with its context, i.e. the sentence it appears in and the preceding and following sentence, as well as various features that could be drawn from the Alpino-parse of the sentence. Of course, only instances where the theme constituent was expressed could be used in the analyses.

Prepositional constituents that were siblings of the verb-node and that were headed by the correct preposition, but that were marked as adjuncts, were nonetheless initially considered as potential prepositional objects. This was done in order to check manually whether they were actual adjuncts, as is the case in (50). If so, they were of course removed from the dataset. This manual checking is described in Chapter 5.

A known issue of the Sonar-corpus is that it contains a number of double sentences. To counter this, only a single sentence was retained if two or more sentences were exactly identical. This extraction was done by a combination of XQuery-, Python- and Bash-scripts (on how to best access XML-treebanks, see Bouma and Kloosterman 2002, 2007 and Augustinus 2015: 99–117).

50. *Naar verwachting zullen ook zakenreizigers    en*
     to   expectation   will   also  business_travelers  and
     *vrachtwagenchauffeurs  bellen.*
     truck_drivers                phone
     'Expectations are that business travelers and truck drivers will call as well.'
                    (WR-P-P-G-0000166907.p.2.s.4)

## 3.4     Levels of abstraction

Before any sort of analysis of the data could take place, it needed to be determined at which levels of abstraction these data could be analyzed. This section describes the procedure used to distinguish these levels, the result of which can be found in Figure 1. We started at the most abstract level of the alternation, i.e. the alternation between the fully abstract transitive and prepositional intransitive construction. Investigating the alternation at this level of abstraction would involve throwing all observations of the various prepositions and verbs in Table 2 into a single heap, i.e. a single dataset, and then trying to analyze that dataset. This would naturally result in a fairly heterogeneous dataset.

We descended one level by keeping the prepositional slot constant. This means that we distinguished several more potential constructions under the fully abstract prepositional intransitive construction, each with its own specific preposition. The preposition-slot was chosen for three reasons: (i) it formed the most obvious difference between both variants of the alternation, since it is the slot that disappears in the transitive construction; (ii) it exhibited the least variation, showing only 16 filler types, i.e. 16 different prepositions (cf. Wible and Tsao 2017: 18–22); and (iii) in previous research, the filler of this slot was typically used to distinguish between several different constructions, such as the English conative construction with the preposition *at* and the English intransitive motion construction with the preposition *into* (Goldberg 1999; Broccias 2001). This latter practice appears to cut ice, because once this slot is kept constant, it becomes easy to discern several semantic groups from the seemingly motley collection of verbs, as presented in Table 2.

Investigating the alternation at the level of the preposition would involve limiting the dataset to only the instances of a single preposition, thereby rendering the data already somewhat more homogeneous. For instance, if we would focus on the preposition *aan* 'on', we would compose a dataset containing only the transitive and prepositional instances of the verbs alternating with that preposition, viz. *bouwen* 'build', *knutselen* 'tinker', *schaven* 'plane', *timmeren* 'carpenter', *gehoorzamen* 'obey', *geloven* 'believe', etc.

The next level was then defined by putting constraints on the verb-slot. This slot was chosen for two reasons. The first is the pivotal role of the verb in theoretical accounts of argument realization. Although this role is downplayed in accounts that employ schematic argument constructions, compared to so-called lexicalist accounts (e.g. Müller 2006; Müller and Wechsler 2014, see Section 1.3), it is generally maintained that its influence is more important than that of the subject- and object-slots (Boas 2008, 2014; Goldberg 2013). The second reason for choosing this slot was that the verb-slot was the remaining open slot that varied least, exhibiting only 101 different filler types. Of course these verbs were selected to

alternate, and this reason is therefore only of secondary importance. Investigating the alternation at this level of abstraction entails limiting the dataset to only the instances of a single verb, e.g. *bouwen* 'build'.

After these choices, two open slots remained, the subject and the object, of which we fixed the object-slot to define the following level on the cline from abstract to concrete. This slot was again chosen for two reasons. First, of the remaining slots, the object is the one that is most directly affected by the alternation, since it may appear either in a nominal or a prepositional constituent. Second, of the remaining slots, it is the one that exhibits the least variation in filler types: the average type-token ratio of the object-slot equals 0.5352, compared to 0.7501 for the subject-slot.[32] Investigating the alternation at the level of the object would mean limiting the dataset to the observations of a single object and a single verb thereby making the data highly heterogeneous and comparable. Of course, as we keep descending the cline from abstract to concrete and our data become ever more homogeneous, they also become ever fewer in number. At this level of abstraction, we expect to be left with only small datasets.

The final level was then defined by fixing the last open slot, the subject. Between each level, one could further distinguish intermediate levels of semantically similar verbs, objects or subjects, such as the motoric verbs, constructive verbs, etc., as is done by Perek (2014) for the English conative alternation, and by us in Table 1. Still, the levels of the preposition, verb, object and subject may be considered as important milestones in the continuum from concrete to abstract.

In order to keep the number of data somewhat feasible for manual checking, the present investigation was limited to the branch of the preposition *naar* 'to'. This branch was chosen because it contains both three seemingly coherent groups of verbs, viz. the telephonic, motoric and venatic verbs, as well as three isolated verbs, viz. *peilen* 'gauge', *verlangen* 'desire' and *zoeken* 'search', and because the excerpt in the *Algemene Nederlandse Spraakkunst* that sparked this investigation, mentions an instance of this branch as its first example (cf. Chapter 1, Haeseryn et al. 1997: 1168).

The following chapter presents our hypotheses. We will have hypotheses relating to the levels of the preposition, the verb and the object. To test these hypotheses, our data were analyzed at each of these levels of abstraction.

---

[32] These ratios were calculated as follows. For every construction at the level of the verb, we calculated the type-token ratios for the subject and the object-slot and averaged over them. When the root of the syntactic head of two subjects or objects were identical, these were regarded as two tokens of the same type. Pronominal realizations were not taken up in the calculations, as they are disproportionally frequent in the subject-slot, and distinct complement clauses were regarded as distinct types. These calculations, as well as all statistical analysis in this thesis were executed with the help of R (R Core Team 2014).

prepositional intransitive construction

NP V PP

Fully abstract level

*naar*-construction

NP V *naar* NP

*aan*-construction

NP V *aan* NP

...

Level of the preposition

motoric verbs

NP V$_{mot}$ *naar* NP

'desire'

NP *verlang naar* NP

'search'

NP *zoek naar* NP

'grab'

NP *grijp naar* NP

...

Level of the verb

'desire time'

NP *verlang naar tijd*

...

'search victim'

NP *zoek naar slachtoffer*

'search word'

NP *zoek naar woord*

...

Level of the object

'rapist search victim'

*verkrachter zoek naar slachtoffer*

'police search victim'

*politie zoek naar slachtoffer*

'rescue worker searches victim'

*reddingswerker zoek naar slachtoffer*

...

Level of the subject

'Rescue workers continue to search for victims'
*Reddingswerkers blijven zoeken naar slachtoffers*
(WR-P-E-G-0000005872.p.67.s.1)

'Rescue workers are still searching victims'
*Reddingswerkers zoeken ook nog altijd naar slachtoffers*
(WS-U-E-A-0000375701.p.1.s.8)

'And rescue workers are still searching for victims of collapsed bridge'
*En reddingswerkers zoeken nog naar slachtoffers van ingestorte brug*
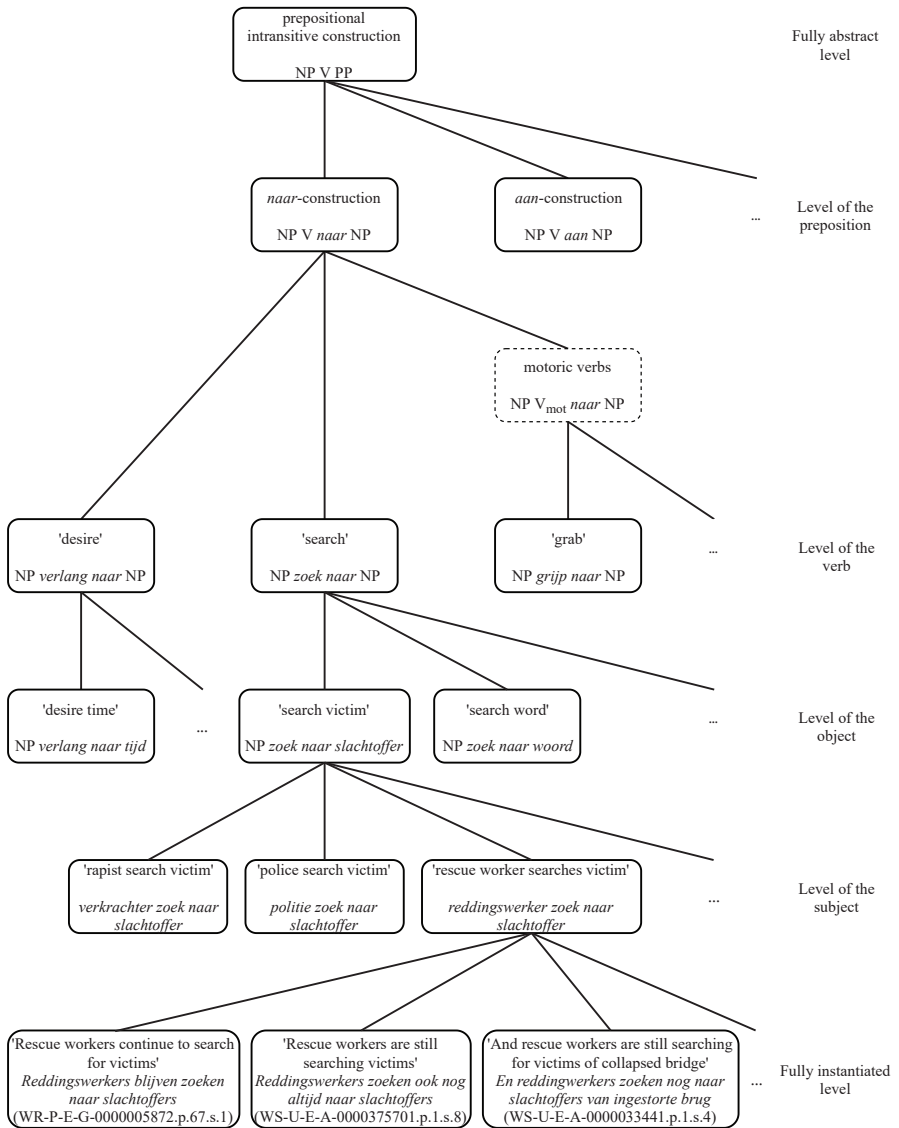(WS-U-E-A-0000033441.p.1.s.4)

...

Fully instantiated level

Figure 1: Possible levels of abstractions at which to investigate the alternation, ranging from entirely schematic to fully concrete.

# 4    Hypotheses

This chapter presents our a priori hypotheses regarding the determinants of the choice between the transitive and prepositional variants.[33] They are divided into the three groups that were introduced in Section 1.2, viz. hypotheses relating to lectal distinctions, semantic differences and processing effects, which will all be discussed in turn. Since the hypotheses relating to processing effects all concern language complexity, we call them the complexity hypotheses.

   For each group, we first present the general reasoning why and how we expect each type of variable to influence variation in argument structure, and discuss earlier research that demonstrated this influence in other case studies. Next, we implement these hypotheses in testable predictions regarding our alternation. Finally, we wrap up this chapter by giving an overview of all predictions.

## 4.1    Lectal Hypotheses

### 4.1.1    Reasoning

As for lectal variables, we will focus on the national distinction between Belgian and Netherlandic Dutch. There are three reasons for this. First, it is arguably the most important binary geographical distinction of the Dutch language as it is spoken in Europe.[34] Second, there is a large body of research available on the differences between both regiolects – although the syntactic differences remain underresearched when compared to the lexical and morphological differences (Haeseryn 2013: 705–710; De Troij et al. subm.) – and we have a good understanding on why and how these differences have developed (Geeraerts, Grondelaers and Speelman 1999: 11–27; Grondelaers, Speelman and Geeraerts 2008). Third, this is the only lectal distinction that our corpus allows us to

---

[33] These are not all hypotheses that will be tested throughout this thesis. Chapter 7 will present a number of additional hypotheses that are based on a data-driven or hypothesis-generating procedure, i.e. that are not a priori.

[34] What might count as a recent indication of this, is that the new *Atlas van de Nederlandse Taal* 'Atlas of the Dutch language' is published in different editions for the Netherlands and Belgium (Jansen et al. 2017; der Gucht et al. 2018).

investigate in any serious depth. The stratification of Sonar into its different components in principle makes it possible to explore some variation in terms of register. However, since we have chosen to exclude the most informal components due to the low quality of their syntactic parses, the remaining bandwidth of register is meager at best (see Section 3.1).[35]

The hypothesized influence of the distinction between Belgian and Netherlandic Dutch on our alternation is rooted in historical sociolinguistics, and to understand how we expect it to influence our alternation, a bit of historical background is required regarding the development of Dutch in the Low Countries. The southern Low Countries counted among the most urbanized regions of northwestern Europe during the late Middle Ages, and the onset of standardization of the local Germanic vernacular began comparatively early (Chandler 1987; van der Wal and van Bree 2008: 102–110, 179–184). This likely happened more or less in tandem with the standardization in the less urbanized North.

This fledgling common development was severed in 1585, when the city of Antwerp fell to forces loyal to Philip II of Spain during the early phases of the Eighty Years' War. The rebel alliance opposing King Philip ultimately managed to hold out in the North, and would go on to form the Netherlandic Republic of the United Provinces, while the southern Low Countries would remain firmly under Habsburg rule (cf. Figure 2; Janssens and Marynissen 2008).

In the North, Dutch became the language of culture and state, and would follow a slow but steady standardization process tightly bound to the development of the Netherlandic nation state and building on the socioculturally dominant Hollandic dialects (Kloeke 1927). Such a gradual evolution is largely mirrored in the standardization processes of the other major European languages (van der Wal 1995; Hüning, Moliner and Vogl 2012). The situation was different in the South. A substantial portion of the Southern socioculturally dominant classes, who were capable of sustaining and promotion standard language planning, had fled to the North in order to escape religious persecution. As a result, the standardization of Southern Dutch was cut short. Consequently, French largely took over the function of supra-regional standard language and the subsequent rulers of the southern Low Countries would passively or actively promote its use as language of culture and state until 1815.

---

[35] Some, though not all, of the material from chat, SMS and Twitter also contains information regarding the sex, age and even residence of the language producers. However, it was decided to exclude this material due to the low quality of its parses (see Section 3.1).
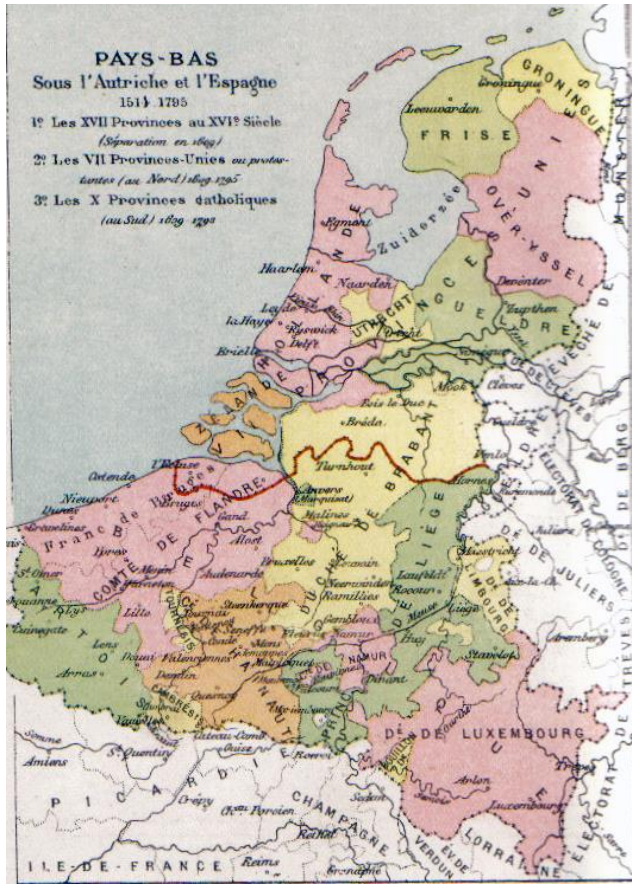
Figure 2: Map of the Low Countries under Habsburg rule, taken over from Gochet (1886: 40). The tick red line indicates the boundary between the South and the newly independent North at the end of the Eighty Years' War. The current state border between Belgium and the Netherlands runs for a large part along this line.

In 1815, the southern Low Countries were joined with the North into the United Kingdom of the Netherlands (Vosters and Weijermars 2011; Vosters and Janssens 2014). The Belgian revolution of 1830, however, ensured that the status of French as the de facto standard language in the South would not be seriously threatened by Dutch – at least not for the time being (Rutten, Vosters and van der Wal 2015).

This only began to change in the final decades of the 19th century, with Dutch being recognized as an official language of Belgium in 1898, in principle on a par with French. Subsequent laws in 1921, 1932 and 1962 divided Belgium into a bilingual area around the capital Brussel and monolingual Dutch-, French- and German-speaking areas. While some proponents of the emancipation of Dutch in

Belgium argued for the creation of a new standard variety of Dutch, endogenic to northern Belgium, others advocated to take over the fully developed Netherlandic standard (Suffeleers 1979; Vosters 2009). The latter faction eventually won out, and the Netherlandic standard was promoted through news media and state usage.

The latter half of the 20th century saw an economic boom of the Dutch-speaking north of Belgium, most notable in the central East-Flemish and Brabantic region between the cities of Ghent, Antwerp, Leuven and Brussels. Here, a new bottom-up process of informal standardization began to gain lift. This new endogenic variety, the so-called *tussentaal* (lit. 'inbetween-language'), has now taken over some of the domains previously dominated by the imported Netherlandic standard, such as informal TV-programs. Still, in more formal domains, the actual Dutch standard language still stands strong, while geographically, the new *tussentaal* has not managed to expand much from its central region of origin (On *tussentaal,* see Geeraerts 2001; Plevoets 2009; Plevoets 2012; Absillis, Jaspers and Van Hoof 2012; De Caluwe et al. 2013; for research specifically focusing on the different types prestige associated with *tussentaal* and standard language, see Grondelaers 2013, Ghyselen 2016 and Rosseel, Speelman and Geeraerts 2018: 23−24 and references cited therein).

To sum up, the status of Dutch in Belgium and the Netherlands may seem deceivingly similar at first sight. In both countries, it is the official language spoken by the majority of the population, where it exists alongside other official languages. During the last five centuries, however, the developments of Netherlandic and Belgian Dutch ran along disparate paths. Dutch in the Netherlands experienced a gradual, relatively natural process of endogenic standardization that resulted in a standard language that is widely spoken in both formal and informal settings. In Belgium, such an endogenic process never really managed to take root until the latter half of the 20th century, which resulted in the *tussentaal* that, however, remained largely confined to the central regions. Instead, the Netherlandic standard was imported, starting from the first half of the 20th century, but its usage was always confined to educated speakers and to formal registers.

This resulted in a situation of Dutch as a pluricentric language consisting of two different national varieties, with Netherlandic Dutch generally being more internally homogeneous than Belgian Dutch (Clyne 1992; De Caluwe 2017).[36] For a characterization of the contemporary status of Standard Dutch in the Netherlands and Belgium, see Grondelaers and van Hout (2010, 2011), Geeraerts (2017) and Grondelaers, van Hout and van Gent (2019); for studies regarding the convergence and divergence of the Belgian and Netherlandic regiolects, see Geeraerts,

---

[36] Further centers of Dutch outside of Europe include Suriname, where Dutch has been the official language since 1876, the Caribbean Netherlands, which consist of Bonaire, Sint Eustatius and Saba, and finally Aruba, Curaçao and Sint Maarten, where Dutch has official status, but is not the lingua franca.

Grondelaers and Speelman (1999) and Daems, Heylen and Geeraerts (2015); for the contemporary status of Dutch dialects in Belgium, see Vandekerckhove (2009) and Ghyselen and Keymeulen (2014); for the contemporary status of Dutch dialects in the Netherlands, see van Hout (1989), Driessen (2005), Goeman and Jongenburger (2009) and the articles in Hinskens and Taeldeman (2013).

## 4.1.2    Predictions

Based on the previous subsection, we formulate the following two hypotheses. First, as Dutch spoken in Belgium tends to be more heterogeneous due to its delayed standardization, we expect its variation to be more difficult to model than Netherlandic Dutch. Hence, we predict lower C-indexes for regression models based on Belgian data compared to those based on Netherlandic data. Such an effect has already been established for the *er*-alternation in presentative sentences with a preposed adjunct, as in (51) (Grondelaers, Speelman and Geeraerts 2002: 344–345; Grondelaers, Speelman and Geeraerts 2008: 189).

51.   *In het redactielokaal   staan (er)     enkele flessen wijn en    wat*
      In  the editorial_room  stand  (there) some   bottles wine and   some
      *borrelhapjes.*
      snacks
      'In the editorial room, there are a couple of bottles of wine and some
      snacks.'
                              (taken over from Grondelaers, Speelman and Geeraerts
                              2008: 158)

Second, related to the first point, we expect the predictors relating to lexical biases, which will be introduced in Section 4.2, to play a more important role in the Netherlands than in Belgium. The reason is that Netherlandic Dutch stands much further in its process of standardization, and what variation remains will more likely have become fixed by lexical constraints or recruited to express a clear meaning difference in a process of functional specialization (Grondelaers, Speelman and Geeraerts 2008: 186).

Similar behavior has also observed for the *er*-alternation, the alternation between the causative auxiliaries *doen* 'do' and *laten* 'let', as in (52), and the alternation between *noemen* 'name/be named' and *heten* 'be named', e.g. (53). The *er*-alternation in the Netherlands was found to be most strongly driven by a distinction temporal and locative adjuncts and the choice of verb, with other factors playing only a minor role at best. Conversely in Belgium, higher level cognitive

factors such as ADJUNCT CONCRETENESS, ADJUNCT TOPICALITY and VERBAL SPECIFICITY needed to be brought into the equation (Grondelaers, Speelman and Geeraerts 2008: 184–185; Grondelaers et al. 2015).

As for the *doen-laten*-alternation, lexical collocations of the auxiliary with the verbs *zien* 'see', *horen* 'hear' and *weten* 'know' play a larger role in the Netherlands than in Belgium (Levshina, Geeraerts and Speelman 2013: 45–46). Finally, the *noemen-heten*-alternation constitutes a straightforward distinction in the Netherlands: *noemen* is only used transitively, meaning 'to call', and *heten* functions as its ergative counterpart 'to be called' (De Grauwe 2014; Speelman 2014: 519–530). In Belgium, however, both *noemen* and *heten* can be used ergatively, and the choice between both forms is more elusive.

52. *De sergeant liet / deed ons door    de modder kruipen.*
    the sergeant let / did  us   through the mud     crawl
    'The sergeant had us crawl through the mud.'
                    (taken over from Verhagen and Kemmer 1997: 62)

53. *Ik  noem  /  heet Dirk.*
    I    name  /  call  Dirk
    'I'm called Dirk.'
                    (taken over from Speelman 2014: 519)

In principle, a third hypothesis could be that register differences will be more outspoken in Belgium than in the Netherlands, but we will refrain from testing this. The reasons are that (i) the register bandwidth of the Sonar-corpus is limited, as mentioned above; but also (ii) more crucially, the distribution of Belgian and Netherlandic material among the corpus components is fairly unbalanced (see Section 3.1). This means that, should we indeed find a difference in the influence of register between both countries, it would hard or even impossible to ascertain whether this is not simply due to the unbalanced distribution.

There is one important caveat concerning our lectal hypotheses, viz. that the author of this thesis is Belgian. Perhaps Belgians are more likely to notice and investigate alternations whose constraints are more elusive in Belgium than in the Netherlands. Moreover, the selection of the alternating verbs described in the previous chapter was done by Belgians, and the selection of alternating instances described in the following chapter was also done by the Belgian author. It cannot be excluded that a Netherlandic researcher might have made different choices in these procedures.

In the selection of the verbs, we have attempted to counter this potential bias by not directly relying on our intuition to come up with potential verbs, but rather

by first generating a list of possibly alternating verbs from the corpus, and then judging each proposed verb based on examples that were again drawn from the corpus (see Section 3.2). Similarly, we have attempted to counter the potential bias in the selection of the instances by basing our decision on the interchangeability of an instance primarily on general linguistic theory and the occurrence of similar instances in the other variant and only secondarily on our own intuitions. When using our own intuitions, we inclined towards a permissive attitude. The only way of fundamentally dealing with this potential caveat however, is simply to encourage Netherlandic researchers to test the same hypotheses on alternations of their choosing.

## 4.2     Semantic Hypotheses

In Section 2.3, we have seen how a wide range of semantic distinctions has already been proposed for the transitive-prepositional alternation. Moreover, similar alternations in argument structure have been argued to be determined by differences in agentivity, volition, concreteness, etc. (Langacker 1991: 359–360; Dowty 1991; Levin and Grafmiller 2012). As a result, the list of potential semantic differences between our variants is sheer endless, especially since agentivity, affectedness, conation etc. are broad semantic notions, that each range over several concrete meaning distinctions that are related to one another, but that are certainly not identical (cf. Section 2.3; for agentivity, see e.g. Dowty 1991; Pijpops and Speelman 2017: 212–220, for conation, see Perek 2015: 105–144; Medina 2017).

    We deal with this proliferation of possibly relevant semantic distinctions in two ways. In the first place, we limit our a priori hypotheses to those semantic differences for which we have an explicit theoretical mechanism of how they may have come into being, just as we did for the lectal hypotheses and as we will do for the complexity hypotheses in Section 4.3. Such a mechanism is described by the lexical origin hypothesis, which will be introduced below. This mechanism will be called the lexical origin mechanism in the remainder of this thesis. In the second place, in Chapter 7, we will employ data-driven analyses to inform hypotheses regarding specific semantic distinctions, that will then be directly tested by manual annotation.

### 4.2.1     Reasoning

The lexical origin mechanism states, in short, that abstract constructions with open slots extract their constructional meaning from the prototypical lexical fillers of

those slots, i.e. from the fillers that are statistically biased to appear more often in the slots (Goldberg 1999; Goldberg 2006; Perek and Lemmens 2010). This constitutes a usage-based mechanism par excellence: the meaning of constructions is not simply an immutable component of the language system, learned and fixed during language acquisition, but is rather constantly being shaped by changing biases in language usage. There is substantial evidence for this mechanism from both corpora and experiments (Clark 1978; Goldberg, Casenhiser and Sethuraman 2004; Casenhiser and Goldberg 2005; Boyd, Gottschalk and Goldberg 2009; Ellis and Ferreira-Junior 2009; Ellis and O'Donnell 2011, 2012; Perek and Goldberg 2015, for an overview and a discussion, see Perek 2015: 79−89).

To exemplify the lexical origin mechanism, let us take a look at the English ditransitive and prepositional dative constructions. Widely simplifying for the purposes of illustration, the lexical origin mechanism would work as follows.[37] The most prototypical lexical filler of the verb slot of the ditransitive construction is the verb *give*, as in (54) (Stefanowitsch and Gries 2003: 227−230). Because the ditransitive construction so often appears with this verb, language users would come to associate the ditransitive construction itself with the meaning of the verb *give*, viz. 'transfer of possession' (Goldberg 1992). In other words, the meaning of *give* would 'rub off' to the argument construction (Perek and Goldberg 2015). As for the prepositional dative construction, the most prototypical lexical filler of its verb slot is the verb *bring*, as in (55) (Goldberg 1999: 202−209; Gries and Stefanowitsch 2004: 206−207). In the same way, the meaning of this verb, viz. 'material transfer', would rub off to the prepositional dative construction (Gries 2012: 482).[38]

54.  *John gave Mary an apple.*
55.  *John brought an apple to Mary.*
                    (taken over from Goldberg 1992: 47)

There are then at least two means of testing the meaning of an argument construction by means of corpus data. The first is to apply the Principle of Semantic

---

[37] For a more accurate, but necessarily more complex description of the constructional meanings of both constructions in English and Dutch and their development, see Gropen et al. (1989), Goldberg (1995: 141−179), Colleman (2009b), Colleman and De Clerck (2009, 2011), Geleyn and Colleman (2015) and Geleyn (2017).

[38] Alternatively, it could be argued that this meaning derives from the verb *put*, which would be the most prototypical filler of the verb slot of the caused-motion construction (Goldberg 1999: 202−209). The prepositional dative construction is often described as a subconstruction of the caused-motion construction, and would inherit (most of) its constructional meaning from this construction (Colleman and De Clerck 2009: 9−12).

Coherence (Goldberg 1995: 50–52; Goldberg 2006: 39–40). This principle states that constructions that are semantically coherent, i.e. that have similar or compatible meanings, will more easily combine with one another (Perek 2015: 24–27). For instance, verbs, i.e. lexical verb constructions, would more often combine with argument constructions that carry a similar meaning.

Coupling this Principle to the lexical origin mechanism may seem a bit circular at first: argument constructions would acquire their meaning from the verbs that often appear in them, and verbs would in turn often appear with argument constructions whose meaning is similar to their own meaning. Such a circular interpretation would constitute a misreading, however. The mechanism could work as follows, and as illustrated in Figure 3 by means of the English dative alternation. The following example is fictitious, and is only meant to illustrate how the lexical origin mechanism and the Principle of Semantic Coherence are not necessarily circular. We certainly do not mean to claim that this is indeed how the English ditransitive construction obtained its meaning.
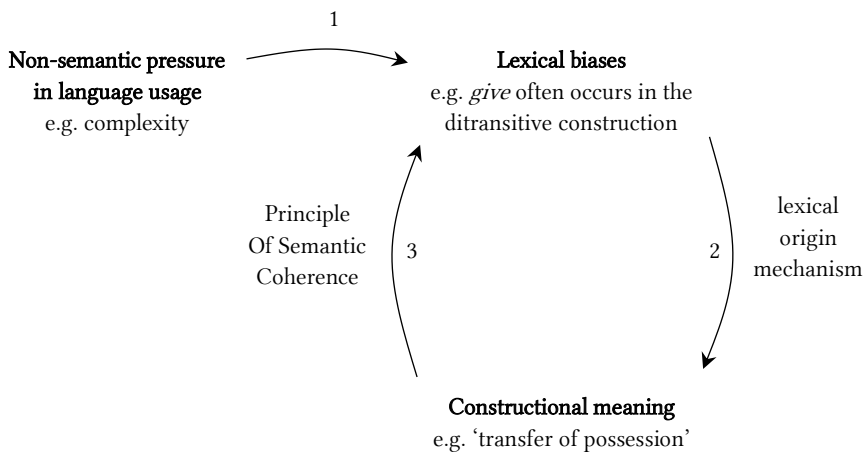


Figure 3: Coupling the lexical origin mechanism with the Principle of Semantic Coherence.

In the first step, a lexical bias emerges in language usage. For instance, perhaps the verbs *give* is often used with simple, informationally light recipients, such as *me* in (56), even more often than other dative verbs. This would cause it to more often appear in the ditransitive construction than those other verbs (Thompson 1990; Bresnan et al. 2007; Bresnan and Ford 2010). Crucially, this ostensible lexical bias does not (yet) have a semantic origin, but is rather caused by something else: in the present example, it is caused by a glut of simple recipients, i.e. by the influence of

language complexity. At this point, we would still expect the ostensible lexical bias of *give* to disappear in a regression model that controls for recipient complexity. The influence of complexity, for that matter, is merely an example of one possible cause; any other mechanism that is known to generate lexical biases could take its place. Such mechanisms include lectal and constructional contamination or simple token frequency (cf. Pijpops and Van de Velde 2016, 2018; Pijpops, De Smet and Van de Velde 2018; Hilpert and Flach forthc.; Lieberman et al. 2007; De Smet and Van de Velde 2019).

56. *Give me your name.*
             (taken over from Bresnan et al. 2007: 84)

In the second step, language users subconsciously notice this lexical bias and come to associate the meaning of *give* with the ditransitive construction, through the lexical origin mechanism. Concretely, language users could (mistakenly) attribute the frequent co-occurrence of *give* and the ditransitive construction to a constructional meaning of 'transfer of possession' of the ditransitive construction. Finally, once language users believe the constructional meaning is in place, they begin to use the construction as such, meaning that the constructional meaning is effectively in place.

In the third step, this would then cause other verbs that are compatible with that meaning to also begin to prefer the ditransitive construction, through the Principle of Semantic Coherence. For instance, perhaps the verb *offer* did not occur particularly often with simple, informationally light recipients, and hence did not originally exhibit an ostensible lexical bias towards the ditransitive construction. Now, however, it does begin to prefer the ditransitive construction since its meaning is highly compatible with the 'transfer of possession' meaning of the ditransitive construction. As a result, we would expect the lexical biases of these verbs to no longer disappear once we control for the original cause, e.g. differences in recipient complexity.

To substantiate the constructional meaning of the argument construction at issue, one would then need to show that (i) the lexical biases of the verbs do not disappear when controlling for e.g. differences in complexity; and (ii) the lexical biases are indeed driven by semantic coherence.

A second means of testing the meaning of an argument construction, apart from using the Principle of Semantic Coherence, is to compare instances of a single verb (Green 1974: 36, 78–146; Gropen et al. 1989; Goldberg 1995: 146–147; Bresnan et al. 2007: 71–74; van Trijp 2015: 626–627). For instance, (57) shows two instances of the verb *send*, one in the ditransitive construction and another in the prepositional dative construction. It would then be argued that the walrus is implied

to end up in Joyce's possession in (57a), while (57b) merely describes that the walrus will move towards Joyce (Langacker 1990: 13–14; Colleman and De Clerck 2009: 9–12). This 'transfer of possession' meaning is not lexically present in (57a). Since the only formal difference between (57a) and (57b) is the choice of argument construction, the meaning difference must have originated there.[39]

57.　a.　*Bill sent Joyce a walrus.*
　　b.　*Bill sent a walrus to Joyce.*
　　　　　　　(taken over from Langacker 1990: 13, cited in Colleman and De Clerck 2009: 10)

Perek and Lemmens (2010) and Perek (2014, 2015: 90–103) discuss problems for the lexical origin mechanism regarding the English conative construction, whose meaning did not appear to be backed by a single or several highly frequent verbs and seemed to differ from one group of verbs to another. The solution proposed by Perek (2014) is to apply the lexical origin mechanism at a lower level, namely the level of verb-class-specific constructions. We can push this reasoning even further. The lexical origin mechanism is fairly general in nature, and contains no elements that are specific to highly abstract argument constructions. As such, we could reasonably expect it to be similarly operative with more concrete argument constructions, even where the verb-slot is fixed. In the next subsection, we apply the lexical origin mechanism to argument constructions at the level of the preposition and the level of the verb, and formulate its predictions at the levels of the preposition, verb and object.

---

[39] This is perhaps most clearly visible in instances of coercion, where language producer – typically consciously – overrides the Principle of Semantic Coherence by combining a verb with an argument construction that starkly contrasts with its meaning (Goldberg 1995: 156–160; Michaelis 2004; van Trijp 2015). For instance, the lexical meaning of the verb *scoop* does not involve a transfer of possession. When combined with a ditransitive construction, however, the meaning of the entire utterance does suddenly involve such a transfer of possession, as in *[N]either Ben nor Jerry ever scooped me an ice cream* (taken from https://www.linkedin.com/pulse/taking-long-view-adam-pekarsky/, accessed September 19, 2019). Meanwhile, when the same verb is used in the prepositional dative construction, as in *[N]either Ben nor Jerry ever scooped an ice cream to me,* the meaning of the entire utterance appears to involve a physical transfer of the ice cream.

## 4.2.2   Predictions

### 4.2.2.1   Level of the preposition

In order to test the lexical origin mechanism at the level of the preposition, we first need to identify the most prototypical verbs that occur in the transitive construction and in the prepositional intransitive construction with *naar* 'to', which we will call the *naar*-construction. To this end, we run a collostructional analysis on the verb slot of the transitive construction and the *naar*-construction (Stefanowitsch and Gries 2003; Gries 2012; Schmid and Küchenhoff 2013; Hilpert 2014b).[40]

We do this separately for the Netherlands and Belgium, because (i) we want to compare our analyses on Belgian data to our analyses on Netherlandic data, as explained in Section 4.1; and (ii) both constructions might be associated with different constructional meanings in the Belgian and Netherlandic regiolects. Table 3 contains the top 5 collexemes ranked according to decreasing collostructional strength, for the transitive construction and the *naar*-construction in the Netherlands.[41] Table 4 contains the same for Belgium. These collexes can be interpreted as the 5 most prototypical fillers of the verb slots of these constructions.

| Transitive construction | | | | *Naar*-construction | | | |
|---|---|---|---|---|---|---|---|
| Verb | Freq. in the transitive | Total freq. | Collostr. strength | Verb | Freq. in the *naar*-cxn | Total freq. | Collostr. strength |
| *hebben* 'have' | 251,812 | 956,878 | 714,608.71 | *gaan* 'go' | 29,525 | 241,201 | 213,367.51 |
| *doen* 'do' | 108,065 | 177,305 | 524,835.64 | *kijken* 'look' | 18,883 | 44,918 | 188,335.39 |
| *krijgen* 'get' | 89,392 | 129,375 | 468,866.61 | *luisteren* 'listen' | 6,034 | 10,868 | 64,252.86 |
| *zien* 'see' | 71,001 | 157,159 | 285,999.55 | *komen* 'come' | 10,717 | 202,317 | 57,636.73 |
| *maken* 'make' | 66,202 | 177,187 | 236,029.88 | *terugkeren* 'return' | 4,351 | 10,943 | 42,352.96 |

Table 3: Top 5 collexemes of the verb slot of the transitive construction and the *naar*-construction in **the Netherlands.**

---

[40] The transitive construction was defined as in Section 3.2, and an occurrence of the *naar*-construction was likewise defined as an occurrence of the prepositional intransitive construction, defined as in Section 3.2., with the preposition *naar* 'to'.

[41] All collostructional analyses were run using publicly available R-code from Gries (2007). The reported collostructional strength measures are based on log-likelihood (Stefanowitsch and Flach 2016: 115–116).

| Transitive construction | | | | Naar-construction | | | |
|---|---|---|---|---|---|---|---|
| Verb | Freq. in the transitive | Total freq. | Collostr. Strength | Verb | Freq. in the naar-cxn | Total freq. | Collostr. strength |
| hebben 'have' | 638,443 | 2,059,105 | 2,055,239.15 | gaan 'go' | 72,097 | 617,020 | 521,661.70 |
| krijgen 'get' | 299,085 | 420,821 | 1,608,613.12 | kijken 'look' | 26,806 | 73,259 | 259,456.84 |
| doen 'do' | 254,415 | 412,182 | 1,252,398.50 | trekken 'pull' | 25,410 | 74,038 | 241,842.42 |
| nemen 'take' | 120,239 | 160,992 | 665,163.60 | verhuizen 'move' | 12,545 | 19,046 | 141,493.51 |
| zien 'see' | 146,687 | 288,184 | 640,225.37 | komen 'come' | 26,276 | 524,224 | 141,250.90 |

Table 4: Top 5 collexemes of the verb slot of the transitive construction and the *naar*-construction in **Belgium.**

The verbs ranking highest in collostructional strength for the *naar*-construction in both Belgium and the Netherlands appear to share a common meaning element of directionality: *gaan* 'go', *trekken* 'pull', *verhuizen* 'move', *komen* 'come' and *terugkeren* 'return' express types of movement, and *kijken* 'look' and *luisteren* 'listen' express the directing of respectively visual and auditory attention at a certain target.

Among the verbs ranking highest in collostructional strength for the transitive construction, such a shared semantic element is less evident. Perhaps the transitive construction, as one of the most basic argument constructions, is not or no longer associated with a particular constructional meaning and merely functions as a formal shell that is semantically empty. Or perhaps it is naïve to expect a single clearly delineated sense for such basic and highly frequent argument construction like the transitive construction (cf. Hopper and Thompson 1980; Dowty 1991; Dodson and Tomasello 1998; Ibbotson et al. 2012). Instead, we could perhaps discern a sense of possession in Table 3 and Table 4 (cf. *hebben* 'have', *krijgen* 'get', *nemen* 'take'), and one of 'acting upon' (cf. *doen* 'do', *maken* 'make', *nemen* 'take').[43]

In Section 4.2.1, we mentioned two ways of using corpus data to test the meaning of argument constructions. The first was based on the Principle of Semantic Coherence, which states that verbs whose meaning is compatible with the meaning of an argument construction would prefer that argument construction. The second way was to keep the verb slot constant, and compare instances of a

---

[43] Similarly, the less frequent ditransitive construction has also been claimed to contain several extended senses (Goldberg 1995: 38, 147–150; Stefanowitsch and Gries 2003).

single verb. This would involve investigating the alternation at the level of the verb. We will return to this second way in Subsection 4.2.2.2.4, and now turn to the first.

Applying the Principle of Semantic Coherence, we predict that verbs that express some form of directionality or movement will more often combine with the *naar*-construction, while those that do not or at least less so, but instead perhaps express a sense of possession or 'acting on', will more often combine with the transitive construction. More concretely, we predict that, in the Netherlands, verbs that are semantically closer to the top 5 collexemes of of the verb slot of the *naar*-construction, viz. *gaan* 'go', *kijken* 'look', *luisteren* 'listen', *komen* 'come' and *terugkeren* 'return' will prefer the prepositional variant. Conversely, verbs that are semantically closer to the top 5 collexemes of the verb slot of the transitive construction, viz. *hebben* 'have', *doen* 'do', *krijgen* 'get', *zien* 'see' and *maken* 'make', will prefer the transitive variant. In Belgium, we likewise predict that verbs that are semantically closer to *gaan* 'go', *kijken* 'look', *trekken* 'pull', *verhuizen* 'move' and *komen* 'come' will prefer the prepositional variant, whereas verbs that are semantically closer to *hebben* 'have', *krijgen* 'get', *doen* 'do', *nemen* 'take' and *zien* 'see' will prefer the transitive variant.

To operationalize that prediction, we need a quantitative measure of semantic coherence. This can be obtained by using distributional vectors, also known as semantic vector spaces or word embeddings (Turney and Pantel 2010). Distributional vectors are becoming increasingly popular in syntactic research (Levshina and Heylen 2014; Perek and Hilpert 2017; Perek 2018; Speelman, Heylen and Grondelaers forth.).[44] They are based on the idea that words that are similar in meaning tend to occur in similar textual contexts. For instance, the verbs *jagen* 'hunt' and *vissen* 'fish' often occur in the context of *garnaal* 'schrimp', while *telefoneren* 'phone' does so less often. Instead, *telefoneren*, 'phone' occurs more often in the context of *nummer* 'number' and *bestuurder* 'driver', just like its synonym *bellen* 'phone'. We can then count for each verb how often they occur in the context of *garnaal* 'shrimp', how often in the context of *nummer* 'number', how often in the context of *bestuurder* 'driver' etc.

This yields for each verb a row of numbers, i.e. a vector, that contains these frequency counts. In the present example, the verbs *jagen* 'hunt', *vissen* 'fish', *telefoneren* 'phone' and *bellen* 'phone' would be called the *target words* of the vector, while *garnaal* 'shrimp', *nummer* 'number', *bestuurder* 'driver' would be called its *context features*. These frequency vectors are then typically weighted in some way to account for the fact that highly frequent context features will often occur with any target words. Finally, we can then calculate distances between the

---

[44] We only talk about type-level distributional vectors here. These are vectors that aim to model the semantics of a single lemma, e.g. the verb *jagen* 'hunt', across multiple occurrences of that lemma. Meanwhile, so-called token-level vectors aim to model the semantics of a single occurrence of a word, and are also increasingly used in linguistic research (Heylen et al. 2015; De Pascale 2019).

vectors. The distance between vectors of target words that are semantically close, will tend to be small, while that between the vectors of target words that are semantically far apart, will tend to be large.

In practice, we calculated dependency-based distributional vectors for all alternating verbs and for all verbs in the top 5's collexemes (Padó and Lapata 2010). The context features of the vectors were based on 7 dependency relations employed in Levshina and Heylen (2014: 31), viz. subject, direct object, indirect object, prepositional complement, locative nominal complement, locative prepositional complement and predicative complement.[45] An example of such a context feature would be *subject:mens/noun*. When building the distributional vector of the verb *zien* 'see', the sentence (58) added 1 to the count of this context feature.

58. *Maar als je    krassen maakt, dan zien de mensen dat   wel.*
    but    if  you  scratch  make   than see  the people  that  PART
    'But if you make scratches, people will notice.'
                (WR-P-E-G-0000000082.p.91.s.1)

Now, of course we expect the top 5 collexemes of verb slot of the transitive construction to generally score high on context features based on direct object relation, and the top collexemes of the verb slot of the *naar*-construction to score high on context features based on prepositional objects or locative prepositional complements with the prepositon *naar* 'to'. Meanwhile, alternating verbs that prefer the transitive construction over the *naar*-construction would also score high on features with direct objects and vice versa. This would introduce a degree of circularity into our models. To prevent this, we blinded the context features of the distributional vectors to our variants. Concretely, the context features do not distinguish between direct objects and prepositional complement or locative prepositional complements with *naar* 'to'. For example, sentence (58) would add 1 to the count of the context feature *direct-object-or-naar:kras/noun* of the verb *maken* 'make', and sentence (59) would add 1 to the count of the context feature *direct-object-or-naar:kist/noun* of the verb *kijken* 'look'.

---

[45] These relations respectively corresponds to the Alpino REL-tags *su, obj1, obj2, pc, ld* with the *cat*-tag not equal to *pp*, ld with the CAT-tag equal to *pp*, and *predc*. We did not build vectors based on subcategorization frames, as the choice of subcategorization frame is exactly what we are trying to predict (Levshina and Heylen 2014: 32–33).

59. *In  de kerk    kijk   je    ook altijd    naar de kist.*
    in  the church  looks  you  also always   to     the coffin
    'In church, you also always look at the coffin.'
                        (WR-P-E-G-0000000013.p.20.s.1)

In calculating the distributional vectors of the verbs, only instances with the POS-tag *verb* and the appropriate ROOT-tag, e.g. *kijk* for *kijken* 'look', were counted. Furthermore, instances of the verbs *hebben* 'have', *krijgen* 'get', *doen* 'do' and *gaan* 'go', where these verbs were used as auxiliaries, were skipped.[46] Context features that correspond to function words, such as *subject:je/pron* for the verb *maken* 'make' in (58) and for the verb *kijken* 'look' in (59), were also disregarded. Only the 5000 most frequent dependency-based features were retained in the vectors as context features, and their co-occurrence frequencies were weighted through positive point-wise mutual information.

Finally, the measure VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION is calculated for each alternating verb as in Equation 1. This measure takes the average cosine similarity of the vector of an alternating verb to the top 5 collexemes of the transitive construction, and subtracts this from its average cosine similarity to top 5 collexemes of the *naar*-construction. The resulting value is then multiplied by 10.[47] For the mathematical definition of the cosine similarity between two vectors, viz. $sim_{cm}(\overrightarrow{verb_1}, \overrightarrow{verb_2})$, and for a comparison to other measures of distributional similarity, see Weeds, Weir and McCarthy (2004).

The higher the measure VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION, the closer the vector of the alternating verb is to the top 5 collexemes of the *naar*-construction and the further away it is from the vectors of top 5 collexemes of the transitive construction. We hence predict this measure to be positively correlated with a proclivity for the prepositional variant. Note that this is a relative rather than an absolute measure of semantic coherence, since it does not calculate the semantic coherence of a verb to the *naar*-construction as such, but rather its semantic coherence to the *naar*-construction relative to the transitive construction.

---

[46] This was done by checking whether these verbs had a verbal complement (Alpino REL-tag *vc*).

[47] Cosine similarities range between 0 and 1, hence differences between them range between -1 and 1. When such a measure is entered into a regression model, it yields conspicuously high estimates. Multiplying the measure itself by 10 simply means these estimates of the regression model will be 10 times lower, but it has no qualitative effect on the regression model.

VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION of the alternating verb $v$

$$= 10(\frac{\sum_{n=1}^{5} sim_{cm}(\overrightarrow{\text{naar\_cxn collexeme}_n}, \vec{v})}{5} - \frac{\sum_{n=1}^{5} sim_{cm}(\overrightarrow{\text{transitive\_cxn collexeme}_n}, \vec{v})}{5})$$

Equation 1: Operationalization of semantic coherence of the verbs.

## 4.2.2.2 Level of the verb

Trying to apply the lexical origin mechanism to argument constructions at the level of the verb may seem odd at first. After all, the verb slot of these constructions is fixed. Any meaning difference between e.g. the transitive *verlang*-construction and the *verlang-naar*-construction cannot have originated in these slots, since they contain the same verb. Still, both constructions do have other open slots. In the previous subsection, we attempted to investigate the meaning of the transitive construction and the *naar*-construction at the level of the preposition by running collostructional analyses on the slot of the next underlying level, i.e. the verb (see Figure 1 in Chapter 3). We will now attempt to get at the meaning of a verb-level argument construction in essentially the same way, by running collostructional analyses on the next underlying slot, i.e. the object slot.

As fillers of the object slot, we look at the theme roots as defined in Section 3.2, viz. the root of the deepest syntactic head of the theme. Only full-nominal theme roots were considered.[48] At this level, we will look at the three isolated verbs in the *naar*-branch (cf. Table 2 and Figure 1 in Chapter 3), viz. *verlangen* 'desire', *peilen* 'gauge' and *zoeken* 'search'.

### 4.2.2.2.1 *Verlangen* 'desire'

The results of the collostructional analyses on the object slots of the transitive *verlang*-construction and the *verlang-naar* construction can be found in Table 5 and Table 6 for respectively the Netherlands and Belgium.

As the most prototypical slot fillers of the object slot of the transitive construction, we find objects such as *tegenprestatie* 'counter effort', *teken* 'sign', *offer* 'sacrifice' and *gehoorzaamheid* 'obedience'. These are objects that are

---

[48] By full-nominal, we mean theme roots with the Alpino CAT-tag *mwu* (multi-word-unit) or the pos-tags *noun, name, adj* (adjective) and *verb*. These last two pertain to participles, such as *overlevende* 'survivor'. In the analyses, we actually used a combination of the ROOT-tag and the POS-tag, in keeping with Levshina and Heylen (2014) and Speelman, Heylen and Grondelaers (forthc.). Only the root tags are shown in the tables presenting the results in order to improve readability.

typically demanded. Conversely, in the object slot of the *naar*-construction, we prototypically find *leven* 'life', *rust* 'rest', *huis* 'home', etc., i.e. objects that are longed for. In this way, we could interpret the results in Table 5 and Table 6 as indicating a difference between a meaning of 'desire' specialized to 'demand' and one of 'desire' specialized to 'long for', as proposed by Broekhuis (2004: 122), Van de Velde (2014a: 340) and den Boon and Geeraerts (2005).

| Transitive *verlang*-construction | | | | Verlang-*naar*-construction | | | |
|---|---|---|---|---|---|---|---|
| Theme root | Freq. in the trans. *verlang*-cxn | Total freq. | Coll. str. | Theme root | Freq. in the *verlang*-*naar*-cxn | Total freq. | Coll. str. |
| *tegenprestatie* 'countereffort' | 7 | 211 | 98.78 | *leven* 'life' | 25 | 61,856 | 141.13 |
| *teken* 'sign' | 11 | 5,654 | 92.63 | *rust* 'rest' | 16 | 8,077 | 140.51 |
| *offer* 'sacrifice' | 7 | 2,081 | 66.56 | *dood* 'death' | 15 | 20,743 | 101.70 |
| *gehoorzaamheid* 'obedience' | 5 | 485 | 58.75 | *huis* 'house' | 14 | 46,079 | 71.10 |
| *helderheid* 'clarity' | 5 | 792 | 53.83 | *kind* 'child' | 16 | 76,597 | 69.73 |

Table 5: Top 5 collexemes of the verb slot of the transitive *verlang*-construction and the *verlang*-*naar*-construction in **the Netherlands.**

| Transitive *verlang*-construction | | | | Verlang-*naar*-construction | | | |
|---|---|---|---|---|---|---|---|
| Theme root | Freq. in the trans. *verlang*-cxn | Total freq. | Coll. str. | Theme root | Freq. in the *verlang*-*naar*-cxn | Total freq. | Coll. str. |
| *service* 'service' | 6 | 4,369 | 61.98 | *kind* 'child' | 23 | 159,621 | 130.55 |
| *creativiteit* 'creativity' | 4 | 4,106 | 38.57 | *rust* 'rest' | 10 | 43,438 | 65.83 |
| *competentie* 'competence' | 3 | 1,403 | 33.63 | *huis* 'house' | 11 | 89,177 | 58.96 |
| *tegenprestatie* 'countereffort' | 2 | 461 | 25.25 | *bed* 'bed' | 7 | 13,695 | 57.12 |
| *offer* 'sacrifice' | 2 | 719 | 23.47 | *tijd* 'time' | 12 | 174,550 | 50.79 |

Table 6: Top 5 collexemes of the verb slot of the transitive *verlang*-construction and the *verlang*-*naar*-construction in **Belgium.**[49]

---

[49] A few instances that occurred only once in the Belgian subcorpus, e.g. *lichtechtheid* ('light realness'), actually scored a higher collostructional strength, viz. 25.50, than *tegenprestatie*

How precisely can such a meaning differentiation have developed through the lexical origin mechanism? In principle, this could have happened in much the same way as sketched for the ditransitive construction above (compare Figure 3 and Figure 4). First, an ostensible lexical bias emerges in language usage. For instance, the prepositional variant occurs more often with the theme *leven* 'life', while the transitive variant occurs more often with *bewijs* 'proof'. Such a bias may be sparked in many ways. For one, *leven* 'life' may more often appear in complex noun phrases that include modifying prepositional phrases and subordinate clauses, as in (60). This could cause the theme *leven* 'life' to more often appear with the prepositional variant than *bewijs* 'proof', as argued in Section 4.3, creating an ostensible lexical bias.

60.　a.　*Joost  verlangt naar het avontuurlijke fantasierijke  leven  dat*
　　　　Joost　desires　to　the　adventurous　imaginative　life　　that
　　　　*hij leidde toen　hij nog  klein  was.*
　　　　he　led　　when　he　still　small　was
　　　　'Joost longs for the adventurous imaginative life when he was still small.'
　　　　　　　　(WR-P-P-H-0000151970.p.37.s.3)

　　　b.　*Men verlangt naar een  gebalanceerd leven  in een  minder*
　　　　one　desires　to　a　　balanced　　life　in a　　less
　　　　*complexe en　gejaagde  samenleving.*
　　　　complex　and　hectic　　society
　　　　'One longs for a balanced life in a less complex and hectic society.'
　　　　　　　　(WR-P-P-G-0000084427.p.2.s.3)

Now, when the verb *verlangen* 'desire' is combined with the object *leven* 'life', the meaning of the sentence generally inclines more towards 'to long for a life' than towards 'to demand a life'. As a result, a tendency would exist in language use where the prepositional variant of *verlangen* 'desire' is more often used in instances whose meaning inclines more towards 'long for' than towards 'demand' − even though strictly speaking, such a semantic difference between both variants is not yet in place. Language users may then subconsciously interpret these tendencies as evidence of an actual meaning difference between both variants. A likely requirement for such a 'probabilistic reanalysis' of usage tendencies to take place,

---

'countereffort' and *offer* 'sacrifice' in the collostructional analysis of the object slot of the transitive *verlang*-construction in Belgium. This high score is a mere coincidental artefact of their status as hapax legomena, however, and we cannot seriously consider these to be prototypical slot fillers. As such, they were skipped.

is that there already exists a functional need for language users to express a distinction between e.g. 'long for' and 'demand'.
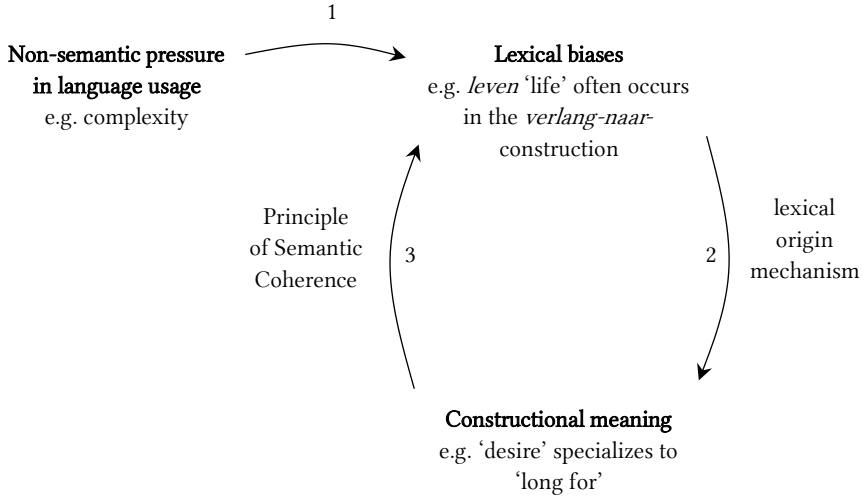


Figure 4: Sketch of one possibility how the meaning of the *verlang-naar*-construction could have developed through the lexical origin mechanism.

Once this has taken place, two things may follow, analogously to what was argued in Section 4.2.1. In the first place, the new meaning distinction may reinforce or alter the lexical biases, such that the biases confirm to the new meaning distinction rather than to their original cause. Put more concretely, objects that are also often longed for, such as *dood* 'death', may develop a preference for the prepositional variant, even though they do not appear in complex noun phrases.

In the second place, the same meaning differentiation could be employed among instances of one and the same theme. For instance, even among instances of the theme root *ding* 'thing', that appears both in the transitive and the prepositional variant of *verlangen* 'desire', the language user would opt for the transitive variant when the meaning of the sentence is akin to 'demanding a thing', and for the prepositional variant when the meaning inclines towards 'longing for a thing'. Note that we again do not claim that this is indeed how the *verlang-naar*- and transitive *verlang*-construction have obtained their meaning. We merely mean to illustrate how the lexical origin mechanism may function for lower-level constructions like the *verlang-naar*-construction, in a way that is analogous to how it functions for higher level constructions.

Our first prediction then pertains to the level of the verb. In particular, we predict that theme roots that are semantically closer to the top 5 collexemes of the *verlang-naar*-construction, of which we suspect that they will be more compatible with the meaning of 'long for', will prefer the prepositional variant, and vice versa. The second prediction pertains to the underlying level of the object, and we will hence come back to it in the Subsection 4.2.2.3.

We will again use distributional vectors to calculate semantic coherence, this time for the themes. Distributional vectors were calculated for all full-nominal theme roots, based on 8 dependency-relations taken over from Levshina and Heylen (2014: 30), viz. subject, direct object, prepositional complement, adverbial prepositional phrase, post-modifying prepositional phrase, adjective, apposition, conjunction. For example, when building the vector of *leven* 'life', the sentence (61) added 1 to the counts of the context features *post-modified-by-prepositional-phrase-with-bij:koraalrif/noun*, *subject-of:loop/verb*, and *modified-by-adjective: veilig/adj*.

61. *Het veilige leven bij het koraalrif loopt ten einde voor de*
    the safe life with the coral_reef runs to_the end for the
    *jongen.*
    younglings
    'The safe life at the coral reef comes to an end for the younglings.'
              (WR-P-E-G-0000000053.p.129.s.1)

Context features with the verb *verlangen* 'desire' were ignored to avoid circularity, as were context features that correspond to function words. Only the 5000 most frequent context features were retained in the vectors, and the frequencies were weighted through positive point-wise mutual information. Finally, SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION was calculated for each theme root as in Equation 2, analogous to Equation 1. Note that this is again a relative measure of coherence. This measure is then predicted to correlate positively with a predilection towards the prepositional variant among the instances of *verlangen* 'desire'.

SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION of theme root $t$

$$= 10\left(\frac{\sum_{n=1}^{5} sim_{cm}(\overrightarrow{verlang\_naar\_cxn\ collex_n}, \vec{t})}{5} - \frac{\sum_{n=1}^{5} sim_{cm}(\overrightarrow{trans\_verlang\_cxn\ collex_n}, \vec{t})}{5}\right)$$

Equation 2: Operationalization of semantic coherence of the theme roots of *verlangen* 'desire'.

### 4.2.2.2.2 *Peilen* 'gauge'

We now turn to the verb *peilen* 'gauge'. The prepositional variant of this verb yields only 13 instances from the Netherlands, so running a collostructional analysis on its object slot based on the Netherlandic data would be pointless. Table 7 contains the results of the collostructional analyses on the object slot of the transitive *peil*-construction and the *peil-naar*-construction in Belgium.

| Transitive *peil*-construction | | | | Peil-naar-construction | | | |
|---|---|---|---|---|---|---|---|
| Theme root | Freq. in the trans. *peil*-cxn | Total freq. | Coll. str. | Theme root | Freq. in the *peil-naar*-cxn | Total freq. | Coll. str. |
| *stemming* 'mood' | 24 | 5,169 | 366.36 | *reactie* 'reaction' | 53 | 25,000 | 541.03 |
| *reactie* 'reaction' | 14 | 25,000 | 153.87 | *tevredenheid* 'satisfaction' | 33 | 2,036 | 470.98 |
| *diepte* 'depth' | 5 | 2,676 | 66.82 | *mening* 'opinion' | 42 | 13,061 | 463.38 |
| *inwoner* 'inhabitant' | 6 | 27,746 | 54.36 | *verwachting* 'expectation' | 40 | 12,421 | 441.37 |
| *mening* 'opinion' | 5 | 13,061 | 50.97 | *kennis* 'knowledge' | 29 | 15,998 | 286.46 |

Table 7: Top 5 collexemes of the verb slot of the transitive construction and the *naar*-construction in **Belgium.**

There are two theme roots that occur both in the top 5 collexemes of the transitive *peil*-construction and the *peil-naar*-constructions, viz. *reactie* 'reaction' and *mening* 'opinion'. If there is indeed a meaning difference between both variants of *peilen* 'gauge', it does seem a lot more subtle than for *verlangen* 'desire'. The non-overlapping top 5 collexemes for the transitive variant are *stemming* 'mood', *diepte* 'depth' and *inwoner* 'inhabitant', and for the prepositional variant *tevredenheid* 'satisfaction', *verwachting* 'expectation' and *kennis* 'knowledge'. We could perhaps say that these last three objects are typically gauged by asking people questions, while the first three involve a more direct way of judging the mood, a depth, or the inhabitants.

We then predict that instances whose theme root is semantically closer to the top collexemes of the transitive *peil*-construction, will prefer the transitive variant. These are the objects of which we suspect that they more typically involve gauging as directly judging. Conversely, theme roots that are semantically closer to the top collexemes of the *peil-naar*-construction, of which we suspect that they are more typically gauged by asking questions, would prefer the prepositional variant. This prediction is operationalized by COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION,

which is calculated as in Equation 3, analogously to Equation 2. Note that only the cosine similarities of the non-overlapping collexemes affect the value of SEMANTIC COHERENCE TO THE PEIL-NAAR-CONSTRUCTION, since the cosine similarities of the overlapping collexemes nicely cancel each other out mathematically, as shown in Equation 3.

SEMANTIC COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION of theme root $t$

$$= 10\left(\frac{\sum_{n=1}^{5} sim_{cm}(\overrightarrow{\text{peil\_naar\_}cxn\ collexeme_n}, \vec{t})}{5} - \frac{\sum_{n=1}^{5} sim_{cm}(\overrightarrow{trans\_\text{peil\_}cxn\ collexeme_n}, \vec{t})}{5}\right)$$

$$= 10\left(\frac{sim_{cm}(\overrightarrow{stemming}, \vec{t})}{5} + \frac{\cancel{sim_{cm}(\overrightarrow{reactie}, \vec{t})}}{5} + \frac{sim_{cm}(\overrightarrow{diepte}, \vec{t})}{5} + \frac{sim_{cm}(\overrightarrow{inwoner}, \vec{t})}{5}\right.$$
$$+ \frac{\cancel{sim_{cm}(\overrightarrow{mening}, \vec{t})}}{5} - \frac{\cancel{sim_{cm}(\overrightarrow{reactie}, \vec{t})}}{5} - \frac{sim_{cm}(\overrightarrow{tevredenheid}, \vec{t})}{5}$$
$$\left. - \frac{\cancel{sim_{cm}(\overrightarrow{mening}, \vec{t})}}{5} - \frac{sim_{cm}(\overrightarrow{verwachting}, \vec{t})}{5} - \frac{sim_{cm}(\overrightarrow{kennis}, \vec{t})}{5}\right)$$

Equation 3: Operationalization of semantic coherence of the theme roots of *peilen* 'gauge'.

#### 4.2.2.2.3    *Zoeken* 'search'

Finally, Table 8 and Table 9 contain the results of the collostructional analyses on the object slots of the transitive *zoek*-construction and the *zoek-naar*-construction, respectively in the Netherlands and in Belgium. As for *peilen* 'gauge', we again find some overlap among the top 5 collexemes. *Oplossing* 'solution' scores first place in all lists except among the collexemes of the transitive *zoek*-construction in the Netherlands, where it ranks fourth.

We look at the other collexemes to try to interpret these results as indicating a meaning difference. One such possible interpretation is the following. The collexemes *toevlucht* 'refuge', *heil* 'salvation', *contact* 'contact', *aansluiting* 'connection', and – perhaps to a lesser extent – *weg* 'road' are objects that only come into being through the act of searching, i.e. are objects that one seeks to make or seeks to acquire. Conversely, the collexemes *overlevende* 'survivor', *oorzaak* 'cause', *alternatief* 'alternative', *manier* 'manner', *mogelijkheid* 'possibility', and to a lesser extent *woord* 'woord', constitute objects that exist independently of the act of searching, i.e. they are objects that one literally looks for. Such an interpretation seems plausible both for the Netherlandic and Belgian results. Again, we predict that theme roots that are semantically close to the top 5 collexemes of each variant will prefer the respective variant, and this prediction is operationalized by the

variable SEMANTIC COHERENCE TO THE *ZOEK-NAAR*-CONSTRUCTION, calculated analogously to Equation 2.

| **Transitive *zoek*-construction** | | | | **_Zoek-naar_-construction** | | | |
|---|---|---|---|---|---|---|---|
| Theme root | Freq. in the trans. *zoek*-cxn | Total freq. | Collostr. strength | Theme root | Freq. in the *zoek-naar*-cxn | Total freq. | Collostr. strength |
| *toevlucht* 'refuge' | 426 | 996 | 6,160.98 | *oplossing* 'solution' | 286 | 9,842 | 3,049.72 |
| *heil* 'salvation' | 405 | 1,457 | 5,427.46 | *woord* 'word' | 143 | 43,330 | 899.47 |
| *contact* 'contact' | 577 | 16,387 | 5,201.29 | *manier* 'manner' | 118 | 30,639 | 777.69 |
| *oplossing* 'solution' | 329 | 9,842 | 2,925.99 | *mogelijkheid* 'possibility' | 99 | 14,225 | 768.40 |
| *aansluiting* 'connection' | 183 | 1,302 | 2,171.12 | *overlevende* 'survivor' | 63 | 1,342 | 730.86 |

Table 8: Top 5 collexemes of the object slot of the transitive *zoek*-construction and the *zoek-naar*-construction in **the Netherlands**.

| **Transitive *zoek*-construction** | | | | **_Zoek-naar_-construction** | | | |
|---|---|---|---|---|---|---|---|
| Theme root | Freq. in the trans. *zoek*-cxn | Total freq. | Collostr. strength | Theme root | Freq. in the *zoek-naar*-cxn | Total freq. | Collostr. strength |
| *oplossing* 'solution' | 1,860 | 31,896 | 18,796.50 | *oplossing* 'solution' | 1,401 | 31,896 | 16,484.53 |
| *contact* 'contact' | 849 | 33,871 | 7,101.39 | *alternatief* 'alternative' | 250 | 11,422 | 2,566.72 |
| *heil* 'salvation' | 509 | 1,621 | 6,988.33 | *manier* 'manner' | 333 | 81,956 | 2,299.70 |
| *toevlucht* 'refuge' | 456 | 1,056 | 6,623.06 | *overlevende* 'survivor' | 155 | 2,432 | 1,927.86 |
| *weg* 'way' | 672 | 60,413 | 4,526.19 | *oorzaak* 'cause' | 202 | 15,492 | 1,862.74 |

Table 9: Top 5 collexemes of the object slot of the transitive *zoek*-construction and the *zoek-naar*-construction in **Belgium**.

#### 4.2.2.2.4    Testing higher-level influence

At the end of subsection 4.2.2.1, we stated our intent to also test the hypothesized meaning difference between the abstract transitive and *naar*-construction at the level of the verb. Concretely, we would expect that among instances of a single verb, those that involve some form of movement or directionality would occur in the *naar*-construction, and those that do not involve such movement or directionality would prefer the transitive construction. Ideally, testing this hypothesis would procede in a manner that makes it comparable to our verb-specific hypotheses at the level of the verb. To this end, we also ran collostructional analyses on the object slots of the abstract transitive construction and the *naar*-construction. The results of these can be found in Table 10 and Table 11 for the Netherlands and Belgium.

The results are highly comparable to the results of the collostructional analyses of the verb slot in Table 3 and Table 4. Among the top 5 collexemes of the objects slot of the transitive construction, it is hard to find a single common denominator. Conversely, the top 5 collexemes of the object slot of the *naar*-construction do have commonality in that they are all places to go to, barring the Netherlandic collexeme *rechter* 'judge' that can still metaphorically function as a place, viz. the court of law. We now predict that, among instances of a single verb, theme roots that are semantically closer to the top 5 collexemes of the *naar*-construction will more often occur in the prepositional variant, whereas those that are semantically closer to the top 5 collexemes of the transitive construction, will occur more often in the transitive variant. This prediction is operationalized by the variable OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION, which is calculated analogously to Equation 2.

| Transitive construction | | | | Naar-construction | | | |
|---|---|---|---|---|---|---|---|
| Theme root | Freq. in the trans. cxn | Total freq. | Collostr. strength | Theme root | Freq. in the *naar*-cxn | Total freq. | Collostr. strength |
| *rol* 'role' | 11,639 | 23,737 | 49,146.76 | *huis* 'house' | 4,795 | 46,079 | 32,219.94 |
| *geld* 'money' | 12,244 | 39,976 | 37,978.72 | *school* 'school' | 1,995 | 30,141 | 11,516.07 |
| *kans* 'chance' | 9,693 | 24,332 | 35,884.49 | *bed* 'bed' | 1,475 | 11,271 | 10,591.28 |
| *indruk* 'impression' | 6,236 | 8,768 | 33,164.76 | *Nederland* 'the Netherlands' | 1,933 | 98,481 | 6,563.171 |
| *gevoel* 'feeling' | 8,299 | 21,525 | 30,074.71 | *rechter* 'judge' | 656 | 15,141 | 3,224.853 |

Table 10: Top 5 collexemes of the object slot of the transitive construction and the *naar*-construction in **the Netherlands**.

| Transitive construction | | | | Naar-construction | | | |
|---|---|---|---|---|---|---|---|
| Theme root | Freq. in the trans. cxn | Total freq. | Collostr. strength | Theme root | Freq. in the *naar*- cxn | Total freq. | Collostr. strength |
| *kans* 'chance' | 45,053 | 105,059 | 176,317.75 | *huis* 'house' | 10,721 | 89,177 | 76,438.97 |
| *rol* 'role' | 22,823 | 41,722 | 103,796.42 | *school* 'school' | 4,978 | 67,362 | 30,398.14 |
| *geld* 'money' | 29,148 | 92,720 | 92,926.83 | *plaats* 'place' | 3,951 | 144,234 | 16,329.82 |
| *werk* 'work' | 33,771 | 141,265 | 88,100.47 | *ziekenhuis* 'hospital' | 2,294 | 36,867 | 13,191.79 |
| *probleem* 'problem' | 27,135 | 113,254 | 70,894.96 | *België* 'Belgium' | 2,895 | 116,692 | 11,406.70 |

Table 11: Top 5 collexemes of the object slot of the transitive construction and the *naar*-construction in **Belgium**.

For the verb *verlangen* 'desire', it is fairly straightforward what this hypothesis would mean in practice. Desiring (to go to) places, as in (62a), would more likely appear in the prepositional variant, whereas desiring things, as in (62b), would more likely appear in the transitive variant. Similarly, for the verb *zoeken* 'search', trying to find a place typically involve going to that place or at least going in the direction of that place, as in (63a). Such instances would then be hypothesized to prefer the prepositional variant. Conversely, instances that do not involve such directionality, or only less so, as in (63b), would be hypothesized to occur more often in the transitive variant. As for *peilen* 'gauge', this hypothesis would mean that places being gauged, as in (64a), would prefer the prepositional variant, while instances where that is not the case, such as (64b), would prefer the transitive variant.

62. a.   *Soms     verlang ik (naar) een witgeschilderde kamer met*
      sometimes desire I (to) a  white_painted room with
      *helemaal   niets   erin.*
      completely  nothing therein
      'Sometimes I desire a completely empty room, painted white.'
            (WR-P-P-G-0000057667.p.32.s.3)

b.  *De koper van pornografie   verlangt (naar) geen zweempje poëzie*
the buyer of   pornography desires to   no   whim   poetry
*om zich      te laten afleiden van zijn 'lauwe    lust' zoals Vladimir*
to himself to let   distract of   his   lukewarm lust as   Vladimir
*Nabokov   het noemde.*
Nabokov   it   called
'The buyer of pornography does not want a whim of poetry to distract
him from his 'lukewarm lust', as Vladimir Nabokov called it.'
(WR-P-P-B-0000000169.p.294.s.10)

63.  a.  *Hugo zocht      ook (naar) de    minder  gekende plekjes.*
Hugo searched  also (to)    the  lesser   known  little_places
'Hugo also searched for the lesser known spots.'
(WR-P-E-G-0000002168.p.253.s.1)

b.  *In laboratoria van cosmeticabedrijven      zoeken vorsers           .*
in laboratories of   cosmetics_companies search  researchers
*(naar) de ultieme  formule om rimpels uit de wereld te  helpen*
(to)    the ultimate formula to  wrinkles out the world  to  help
'In laboratories, researchers are searching for the ultimate formula to
rid the world of wrinkles.'
(WR-P-P-G-0000184641.p.2.s.1)

64.  a.  *Met een zelden gezien mededogen peilt  Chabrol hier (naar) de*
with a    rarely  seen  compassion gauges Chabrol here (to)    the
*diepste  afgronden van  la  nature humaine.*
deepest  abysses  of   la  nature humaine
'With a rarely seen compassion, Chabrol gauges the deepest abysses of
human nature.'
(WR-P-P-H-0000114862.p.4.s.4)

b.  *Dertig van  de kandidaten worden uitgenodigd voor een  interview,*
thirty of   the candidates  are     invited      for  an   interview
*en   dan peilen we (naar) hun  motivatie.*
and then gauge  we (to)    their motivation
'Thirty candidates are invited for an interview, and then we gauge
their motivation.'
(WR-P-P-H-0000070932.p.8.s.4)

It is perfectly possible that these predictions are confirmed, while the prediction of
the same hypothesis at the level of the preposition, viz. that verbs which express
directionality more often occur in the *naar*-construction, is not. Such a result would

imply that the degree of directionality required from a sentence to opt for the *naar*-construction, is calibrated separately for each verb.

For example, perhaps (63b) expresses more of a directional operation than (62a). After all, (63b) at least expresses an attempt to get to the formula, while (62a) expresses no such attempt to get to the room. As such, we could expect the prepositional variant to be used in (63b) and the transitive variant to be used in (62a). This would be in line with the prediction at the level of the preposition, which stated that verbs that express more directionality, such as *zoeken* 'search', would have a general preference for the *naar*-construction compared to verbs that express less directionality, such as *verlangen* 'desire'. It could also be reasoned, however, that (62a) is particularly directional for an instance of *verlangen* 'desire', while (63b) is rather stationary for an instance of *zoeken* 'search'. Therefore, the prepositional variant should be predicted for (62a) and the transitive variant for (63b). This would correspond to the prediction at the level of the verb.

In fact, Pijpops and Speelman (2017) report such a finding for the Dutch psych verb alternation, as in (65). This alternation was hypothesized to be influenced by a semantic difference in agentivity. In particular, it was predicted (i) that verbs whose lexical meaning implied a more agentive experiencer would more often appear in the experiencer-subject reflexive construction, and (ii) that individual instances with a more agentive stimulus would exhibit an increased probability of the stimulus-subject transitive construction. The second prediction was confirmed, while the first was not.

65.  a.  *Elizabeth ergert  John.*
      Elizabeth  annoys   John
      'Elizabeth annoys John.'
                    (Transitive construction, stimulus-subject; taken over from
                    Pijpops and Speelman 2017: 210–211)

     a.  *John ergert  zich     aan Elizabeth.*
      John annoys   himself  on    Elizabeth
      'Elizabeth annoys John.'
                    (Reflexive construction, experiencer-subject)

### 4.2.2.3    Level of the object

When we applied the lexical origin mechanism at the level of the preposition, we hypothesized a meaning distinction in terms of directionality. This allowed us to

formulate a prediction at the level of the preposition, and at the next underlying level, viz. the level of the verb. In addition, we also applied the lexical origin mechanism directly at the level of the verb, and hypothesized three verb-specific meaning distinctions. We now also want to test these verb-specific distinctions at the next underlying level, viz. the level of the object. In practice, this means comparing instances that have both the same verb and the same object.

Of course, fixing both the verb slot and the object slot severely restricts the amount of observations in a dataset. For the relatively infrequent verbs *verlangen* 'desire' and *peilen* 'gauge', the shortage of data becomes particularly critical. In fact, for *verlangen* 'desire', the only full-nominal theme roots whose least frequent variant occurs more than once in both Netherlandic and the Belgian data are *ding* 'thing' and *tijd* 'time'. We will hence take these theme roots under scrutiny. Our prediction is then that if the agent demands things or demands time, the transitive variant will be used, whereas when the agent longs for things or longs for time, the prepositional variant will be used.

For *ding* 'thing', we see no way of operationalizing this distinction in a straightforward manner, so we will attempt to code for it directly. For *tijd* 'time', we will distinguish between instances where *tijd* 'time' is a mass noun, and those where it is a count noun. The underlying idea is that people may demand more time to perform a certain task, which involves time as a mass noun, while they long for some time period in the past or present, i.e. they long for time as a count noun. The distinction between count and mass noun is preferred over directly annotating for demand vs. long for, since both labels are more sharply delineated from each other.

For the verb *peilen* 'gauge', we will look at the objects *stemming* 'mood' and *reactie* 'reaction'. These are the only full-nominal objects whose least frequent variant occurred more than five times in the Belgian data.[50] Here we predict that when it is clear from the context that the act of gauging occurred by asking questions, the prepositional variant will be used, while when it is clear from the context that this wasn't the case, the transitive variant will be used. We see no simple way of operationalizing this difference through a more clear-cut distinction, so we will attempt to annotate for it directly.

The highly frequent verb *zoeken* 'search' sports a large number of objects that regularly occur in both variants. As such, we choose two objects where the – at times somewhat elusive – hypothesized distinction between 'seek to make/acquire' and 'look for' can be made in a clear-cut way. The first is *slachtoffer* 'victim'. The meaning 'seeking to make victims' or 'causing victims to come into being' only makes sense if the agent performing the search is an aggressor. Conversely, cases of 'looking for victims' or 'trying to find a number of pre-existing victims' can also

---

[50] No prediction is formulated for the situation in the Netherlands concerning *peilen* 'gauge', as the prepositional variant was nearly non-existent there (see above).

involve an agent that intends to help to victims. As such, we predict that if the agent is an aggressor, the transitive variant will be used, whereas if the agent is a helper, the prepositional variant will be preferred.

The second object is *woord* 'word'. The meaning 'seeking to make words' or 'causing words to come into being' can be interpreted to involve an agent trying to utter something, or trying to build a sentence. Conversely, instances of 'looking for words' or 'trying to find some pre-existing word' would involve, say, an agent looking for words in a certain text. To make this distinction in a straight-forward manner, we will discriminate between those instances where the word(s) in question are unspecific or *de dicto*, and those observations where the word(s) are specific or *de re*. In the first case, any of several words may do, given that they meet particular requirements, e.g. express the correct idea or concept, while the second case would more likely involve looking for multiple pre-determined words in a body of text. We then predict that in the first case, the transitive variant would be used, and in the second, the prepositional variant would be preferred.

### 4.2.2.4        Remaining questions

One could wonder why, given that we are interested in meaning differences between both variants, we did not choose to perform distinctive collexeme analyses rather than collostructional analyses (Gries and Stefanowitsch 2004; Colleman 2009a). Distinctive collexeme analyses do not yield the prototypical fillers of the slot of a construction as such, but rather the lexical items or collexemes that maximally distinguish one construction from another construction. Put concretely, a distinctive collexeme analysis on the object slot of the transitive *zoek*-construction and the prepositional *zoek-naar* construction would have yielded a list of theme roots that characterize the transitive *zoek*-construction **as opposed to** the *zoek-naar*-construction and vice versa.

There are three reasons for the choice of 'ordinary' collostructional analyses over distinctive collexeme analysis. First and foremost, the lexical origin mechanism states that constructions obtain their constructional meaning from the most prototypical fillers of their slots as such, not from the most prototypical fillers of their slots as compared to some other construction. Since we want to test predictions based on this mechanism, we need to stay as close to its formulation as possible.

Second, distinctive collexeme analyses would highlight any differences between both constructions, even if they are minor or nearly non-existent. For instance, for the verb *zoeken* 'search', the theme root *oplossing* 'solution' would be assigned a single score that indicates to what degree it is attracted to the one variant and repulsed by the other, even though it is a prototypical filler of the object slot of both

constructions. In this way, distinctive collexeme analyses could lead us astray by yielding overly distinctive and hence optimistic results.

Third, related to this second point, selecting the top 5 'ordinary' collexemes rather than the top 5 distinctive collexemes constitutes a conservative research design. That is, we are essentially making it harder on ourselves by not highlighting any possible difference between both variants.

It could also be asked why we chose to perform the collostructional analyses before the manual selection of the data in Chapter 5, rather than after it. There are four reasons for this. First, we want to have formulated our a priori hypotheses prior to seeing the data that will be used to test these hypotheses. Second, our goal is to test hypotheses regarding the semantics of the constructions. As such, we cannot use semantic criteria to define the constructions when formulating these predictions, as that would render the predictions circular. The constructions hence need to be defined on purely formal grounds. Third, the datasets in their current state will contain more noise than the datasets after Chapter 5, and hence this again constitutes a conservative design choice. Fourth, the collostructional analyses on the verb and object slots of the abstract transitive construction and the *naar*-construction necessarily needed to occur in this way, as it is practically impossible to manually check each occurrence of every verb that appears in the transitive construction and the *naar*-construction in the Sonar-corpus, and we want to keep these analyses as analogous as possible to the collostructional analyses on the object slots of the transitive *verlang*-construction, the *verlang-naar*-construction, the transitive *peil*-construction etc.

## 4.3 Complexity Hypotheses

Concerning the influence of language processing on our alternation, we will focus on hypotheses that have been argued to underlie the Complexity Principle of Rohdenburg (Rohdenburg 1996, 2016). The current section is a slightly reworked version of the text published in Pijpops et al. (2018). We first introduce the Complexity Principle and present a short overview of the various mechanisms that have been proposed to explain this Principle. Next, we apply these mechanisms to our alternation and explain how they make different predictions regarding the variation at issue.

## 4.3.1      Reasoning

The likelihood with which language users insert optional words or morphemes that explicitly mark syntactic structure, tends to increase in complex grammatical environments. This positive correlation between explicitness and complexity is most famously expressed in Rohdenburg's Complexity Principle:

> *In case of more or less explicit grammatical options, the more explicit one(s) will tend to be favored in cognitively more complex environments.* (Rohdenburg 1996: 151)

This correlation has been observed for a multitude of case studies in both naturally occurring language and experimental settings (for an overview, see Jaeger 2010: 23–27). For example, Thompson (1990: 249) was one of the first to show that, as the recipient in a dative clause (e.g. *Nim* in (66)) becomes longer and thus more complex, it was more likely to be introduced by the preposition *to*. Numerous other studies have since then confirmed this finding for both the English and the Dutch dative alternations (e.g. Bresnan et al. 2007; Theijssen 2012; Geleyn 2017; Röthlisberger 2018a; Dubois forthc. and references cited therein). Likewise, Bouma and Kloosterman (2007) and Van Beveren, Colleman and De Sutter (2019) demonstrate that as Dutch sentences with an infinitival complement clauses, e.g. (67), become more complex, the probability of the optional complementizer *om* increases.

66.    a.   *Laura gave Nim a bagel.*
       b.   *Laura gave a bagel to Nim.*
                  (taken over from Thompson 1990: 239)

67.   *De Indiërs   aarzelen   (om) te   investeren in   Uganda.*
     the Indians   hesitated   (to) to invest      in   Ugunda
     'The Indians hesitated to invest in Uganda.'
                  (taken over from Bouma 2017: 56)

Most of the research on this topic, however, has looked into the English *that*-alternation, as in (68) (a.o. Bolinger, 1972; Ferreira & Dell, 2000; Ferreira & Hudson, 2011; Ferreira & Schotter, 2013; Jaeger, 2005, 2010, 2011; Jaeger & Wasow, 2005; Roland et al., 2006). There is also ample evidence for the Complexity Principle for this alternation. The probability of the appearance of *that* has been shown to positively correlate with the length of the subject of the complement clause, as well

as with other operationalizations of complexity (for an overview, see Shank et al., 2016b: 202–213).

68.    *I would guess (that) Al Gore will not endorse anyone.*
                    (COCA, cited in Shank et al., 2016b: 208)

Still, while the correlation described by the Complexity Principle has generally been accepted, there is still disagreement on why it exactly holds true. The different explanations of its cause can be divided into three viewpoints. The first viewpoint asserts that the Complexity Principle is chiefly caused by cognitive processing during language production (e.g. Ferreira & Dell, 2000; MacDonald, 2013). Explicit coding would present a convenient way to buy time for the language producer when processing demands are high, such as in complex linguistic environments.

The second viewpoint states that the Complexity Principle is fundamentally the result of restrictions on the physical language channel (e.g. Fenk-Oczlon, 2001; Fenk & Fenk-Oczlon, 1993; Jaeger, 2010). These restrictions introduce noise into the language channel that may disrupt the flow of information, and as a result, additional coding is required to smooth out the peaks in information density that typically arise in complex environments.

The third viewpoint proposes that the correlation emerges primarily due to cognitive comprehension processing (e.g. Bolinger, 1980; Clark & Murphy, 1982; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Hawkins, 2004). That is, explicit coding is first and foremost aimed at optimizing the addressee's comfort. More complex environments would then be coded using the more explicit grammatical option, because the explicit coding of the syntactic structure simplifies parsing. In what follows, these three viewpoints are introduced in more depth.

### 4.3.1.1    Production viewpoint

The most direct way in which complexity can affect explicitness is through cognitive production processing. Making sentence structure explicit by including the optional complementizer *that* or the preposition *naar* 'to' evidently requires some effort from the producer, but this effort would buy time for the producer to formulate a complex complement clause or noun phrase, thereby relieving pressure on production facilities (Ferreira and Dell 2000: 298–300). The primary cause of the correlation between complexity and explicitness would then be the cognitive effort of the producer.

It is still possible that the comprehender also benefits from the use of explicit coding in complex contexts, but only in a derived or secondary way. Two

production accounts that allow for this are the PDC-model proposed in Gennari & Macdonald (2009), MacDonald (2013), MacDonald & Thornton (2009) and the 'collateral signals' account (cf. Clark, 2004: 373–381, as well as Brennan & Williams, 1995; Clark & Fox Tree, 2002; Collard, Corley, MacGregor, & David, 2008; Corley & Hartsuiker, 2003; Fox Tree & Clark, 1997; Smith & Clark, 1993 and references cited therein).

PDC stand for Production-Distribution-Comprehension. According to the PDC-model, pressures in **P**roduction processing determine the distributions that we find in language use. In turn, these **D**istributions shape an individual's grammar, and finally, this probabilistic grammar is employed in **C**omprehension. This means that the comprehender will expect the form of new sentences to confirm to this grammar, and thus to the form of previously heard sentences, whose realization was optimized for production. The fact that the form of all sentences would consistently be optimized for production would in that way indirectly help the comprehender. When a newly heard sentence would contradict the comprehender's expectations by not being optimized for production, but rather for comprehension, this would – seemingly paradoxically – cause comprehension difficulties.

According to the collateral signals account, the presence of optional markers in explicit coding informs the comprehender about the state of production. For example, the use of such markers may indicate ongoing production difficulties, that may in turn be a cue to the comprehender that the following words are difficult to integrate in the existing context. The comprehender can then prepare for this by cancelling his or her expectations about upcoming material (Grondelaers et al. 2009: 159–160).

### 4.3.1.2    Channel viewpoint

The channel perspective is rooted in Shannon Information Theory (Shannon 1948; Cover and Thomas 1991). It searches the root cause of the Complexity Principle not in any kind of cognitive processing done by either producer or comprehender, but rather in the physical language channel between producer and comprehender (Fenk and Fenk-Oczlon 1993; Fenk-Oczlon 2001; Levy and Jaeger 2007; Jaeger 2010). As such, it is different from both the production and comprehension perspective.

This perspective states that human language use constitutes a form of information exchange, and the language channel is a type of information channel (cf. Coupé et al. 2019). Like any kind of physically existing information channel, the language channel is prone to noise. This noise introduces the risk of information loss. The more information is packed into a signal, e.g. into a string of words, the more information will be lost if the signal is damaged by noise. In other

words, the more dense the information flowing through a channel, the higher the risk of noise causing substantial information loss. Meanwhile, the less dense the information flowing through the channel, the less efficiently the channel is being used. As a result of these competing pressures, any information channel has an associated optimal degree of information density that balances risk of information loss with efficiency of use. The users of a channel will attempt to approximate this level at all times, resulting in a more or less constant density of the information flow through the channel. This has been called the principle of Uniform Information Density (Jaeger 2010).

The channel of natural language has been noted to be particularly prone to noise (Levinson 2000: 28). For example, in the case of spoken language, background noises may cause some words to become unrecognizable to the comprehender. If the producer then chooses to express his message in as few words as possible, such noises may already cause too much information to be lost and may thus render the original message irretrievable. In the case of written language, sources of noise include typing errors, imperfect eyesight, bad printing quality and illegible handwriting. In the case of sign language, they may include sore muscles and visual clutter.

Optional markers that make syntactic structure explicit, such as English *that* or Dutch *naar* 'to', may then present a way to tune the information density of an utterance. Such markers will be low in inherent information content, as they can apparently be added or removed without drastically altering the message expressed by the sentence. Additionally, they explicitly flag what follows as respectively a complement clause or a theme argument, hence rendering it more predictable. According to Information Theory, information equates with the negative logarithm of predictability. This means that by rendering what follows more predictable, these markers effectively reduce the information density of the following complement clause or theme argument. As a result, since complex elements tend to be high in information density and simple elements tend to be low, these markers would be preferable placed in front of complex elements, and omitted in front of simple elements, such that the information density would remain more of less constant. This would then result in the correlation described by the Complexity Principle (Jaeger 2010: 26–28).

In this text, we present the channel-driven account separately from both the comprehension and production perspective for two reasons. First and foremost, it is fundamentally different from both the comprehension and production perspective in stating that the root cause of the Complexity Principle is not to be sought in any kind of cognitive processing, but rather in the physical limitations on the language channel. Second, if one would have to include it under either the production or comprehension perspective, it is not clear which one would be more appropriate. On the one hand, the channel-driven account pivots on successful communication. The question is whether the information contained in the message

reaches the comprehender, and one could therefore include it under the comprehension perspective (cf. Jaeger 2013). On the other hand, the noise in the language channel and therefore the cause of maximal information density stems for a large part, though not completely, from properties of the producer, namely the limitations of our physical articulators (Levinson 2000: 28). Moreover, Ferreira & Schotter (2013: 1569) have argued for a strong affinity between the channel- and production-driven accounts, viewing them as merely *"different levels of description of the same sort of phenomenon"*. According to this viewpoint, the production-driven account would be seen as the cognitive implementation of the principle of Uniform Information Density, which makes sure that language producers in practice always approximate the optimal degree of information density.

### 4.3.1.3    Comprehension viewpoint

Finally, explicitly encoding the syntactic structure of a sentence evidently simplifies parsing and thus comprehension. In the case of *that*, the optional marker signals to the comprehender that the producer is entering a complement clause. In the case of *naar*, the optional word is a preposition explicitly linking the verb to its complement (Haeseryn et al. 1997: 1168–1179; van der Horst and Van de Velde 2008).

   Still, the choice whether or not to use such optional elements of course rests with the producer, not the comprehender. There are then two ways in which comprehension processing can still affect this choice. The first is speaker's altruism or strong audience-design (Kirby 1999: 60; Hawkins 2002; Hawkins 2004). It states that, if the producer is going to utter a complex phrase, s/he will choose the structure that is easiest to parse for the comprehender, even if this requires more effort from his/her part. Of course, the producer then needs to have some way of knowing which structure is easiest to parse, i.e. s/he needs to have access to some metric of parsing effort.

   Note that this account of speaker's altruism is not a case of true altruism, as the producer may also indirectly benefit from forming easily comprehensible sentences. For one, comprehenders may be more inclined to listen to and act on the messages formulated by such producers. Moreover, communication is fundamentally a collaborative task, meaning that producers have to make at least some effort in order to be comprehensible (Zipf 1949).[51] It then only seems a minor step to say that they also make the effort of using optional markers in order to be **easily** comprehensible.

---

[51] Unless, of course, when the producer is consciously trying not be comprehensible, in which case communication ceases to be a collaborative task. However, it is generally assumed that such situations present the exception, rather than the rule (Clark 1996).

The second way in which comprehension processing may affect choices in production is hearer selection (Kirby, 1999: 31–62, see Ferreira & Schotter, 2013: 1568 for a similar proposal). This differs from speaker's altruism in that comprehension steers production in a more indirect way. It proposes that only constructions that lead to successful comprehension become entrenched in grammar, or that those that lead to more effortless comprehension become more strongly entrenched than those which require more effort. Once entrenched in grammar, these constructions can in turn affect the production of the language user in question. In other words, tendencies that obstruct comprehension processing are selected against in language evolution. While this account dispenses with the assumption that some metric of parsing effort is directly taken into account during production, it does require that entrenchment be dependent on successful or easy comprehension. This proposal can be seen as the reversal of the PDC-model from the production perspective. Figure 5 presents a comparison of both.
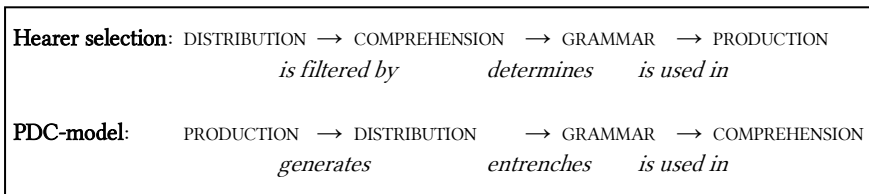
| **Hearer selection**: | DISTRIBUTION | → | COMPREHENSION | → | GRAMMAR | → | PRODUCTION |
|---|---|---|---|---|---|---|---|
| | | *is filtered by* | | *determines* | | *is used in* | |
| **PDC-model**: | PRODUCTION | → | DISTRIBUTION | → | GRAMMAR | → | COMPREHENSION |
| | | *generates* | | *entrenches* | | *is used in* | |

Figure 5: hearer selection versus the PDC-model.[52]

So far, empirical findings from experiments and corpora appear to favor the channel and production perspectives over the comprehension perspective. Ferreira & Dell (2000) find no evidence that language users employ explicitness to simplify comprehension in controlled experiments, while they do find evidence that lexical availability during production plays a role. Likewise, Elsness (1984) and Roland et al. (2006) find no indications that, in corpora, language users use the optional complementizer *that* to facilitate comprehension processing. Further indications from experiments and corpora in favor of the channel perspective are presented in Fenk-Oczlon (2001), Jaeger (2010), Levy & Jaeger (2007); in favor of the production perspective in Ferreira & Hudson (2011), Ferreira & Schotter (2013), Gennari & Macdonald (2009), Kraljic & Brennan (2005), MacDonald (2013), and MacDonald & Thornton (2009). For other studies investigating the differences between

---

[52] DISTRIBUTION here refers to natural language usage as we find it in corpora, and GRAMMAR stands for the cognitive organization of one's experience with language (Cf. Section 2.1, Bybee 2006: 711).

cognitive processing in language production and comprehension, see Bock, Irwin, & Davidson (2004), Tanner & Bulkes (2015), Tanner, Nicol, & Brehm (2014), and references cited therein.

### 4.3.2    Predictions

There are at least two principle ways of contrasting the three perspectives using corpus data. The first is to formulate three separate operationalizations of complexity, each tailored to each perspective. For instance, one operationalization would be more suited for production complexity, while another operationalization would better measure information density, etc. (see Menn and Duffield 2014 for a discussion on several operationalizations of complexity). Next, we could investigate which is the best correlate of explicitness, viz. in our case, the best correlate of the probability of the prepositional variant. However, these various operationalizations of complexity are likely to strongly correlate with one another, making it hard to disentangle them.

We therefore opt for the second way, which is to employ a single operationalization of complexity that works for all perspectives. We can then compare the contexts in which it is correlated with explicitness. In what follows, we first introduce this operationalization of complexity, and then explain how the three viewpoints make different predictions regarding the contexts in which this operationalization should be positively correlated with explicitness. Finally, we discuss whether our data are suitable to test these predictions.

### 4.3.2.1    Operationalization of complexity

As an operationalization of complexity, we use the variable THEME COMPLEXITY, counted as the natural logarithm of the number of words of the theme argument. While this may not constitute the most advanced operationalization of complexity, it is robust, reliable, and largely independent of the employed parsing formalism. Still, it is dependent on the delineation of the theme argument, or in other words, on the question where to draw immediate constituent borders. We assume that these are largely unproblematic though, in that most current linguistic theories would by and large agree on them. In delineating the theme argument, we start from the constituent borders assigned by the Alpino-parser, which is based on HPSG theory (van Noord 2006; van Noord et al. 2013). These borders will be manually checked in the Chapter 5, where we will resume this point in Section 5.2.

For the production and comprehension perspective, the choice for THEME COMPLEXITY is quite straightforward. Regarding the production perspective, the optional preposition *naar* always appears right in front of the theme argument, at

exactly the moment when the producer needs to plan the theme. When the theme is long and the producer is hence under a lot of processing pressure, this would be the most opportune moment to buy extra processing time. Regarding the comprehension perspective, having to parse a long noun phrase puts a large strain on cognitive comprehension facilities. As such, it would be most useful to have a formal marker right in front of this noun phrase that explicitly marks it as the theme argument.

The operationalization of THEME COMPLEXITY for the channel perspective requires some more clarification. It is based on the assumption that longer themes tend to be more specific than short themes. In turn, more specific themes are harder to predict, which means they contain more information. For instance, the theme in (69a) is a lot more specific and harder to predict and hence contains more information than the theme in (69b). As argued above, themes that contain more information have a greater need for a preceding preposition *naar* 'to' to reduce their information density.

69.  a.   *En   hij  peilt   (naar) de  impact van  de  gemeenteraadsverkiezingen*
          and  he  gauges (to)    the impact of    the municipal_elections
          *op de agenda  van  de  federale regering.*
          on the agenda  of    the federal   government
          'And he gauges the impact of the municipal elections on the agenda of the federal government.'
                    (WR-P-E-C-0000010033.p.1.s.4)

     b.   *De BBC peilde   (naar) wat    reacties.*
          the BBC gauged  (to)     some reactions
          'The BBC gauged a number of reactions.'
                    (WR-P-P-G-0000143874.p.3.s.2)

## 4.3.2.2    Differentiating between the three viewpoints

In order to differentiate between the three perspectives, we will distinguish between those instances where the verb precedes the theme as in (70a), and those where the theme precedes the verb as in (70b). Instances where the initial part of the theme precedes the verb and the remainder follows it, as in (70c), are counted amongst those where the theme precedes the verb, since the preposition *naar*, if it is present, would also precede the verb and so does the syntactic head of the theme. In such instances, the part of the theme following the verb is also taken up in the calculation of THEME COMPLEXITY.

We will now argue that the production and channel perspectives predict a negative correlation between the complexity of the theme and the propensity for the explicit prepositional variant in cases such as (70b-c) and positive correlation in cases such as (70a). Meanwhile, the comprehension perspective will be argued to predict a positive correlation in both cases, and even a stronger positive correlation in cases such as (70b-c) than in cases such as (70a).

70.  a.  *We zoeken naar de oorzaak, maar hebben nog geen idee.*
         we search to the cause but have still no idea
         'We are searching the cause, but we have no idea so far'
               (WR-P-P-G-0000039610.p.2.s.5)

     b.  *We zijn dus wel gedwongen nu al       naar een goede*
         We are thus part forced     now already to a    good
         *vervanger te zoeken.*
         substitute to search
         'We are thus forced to already search a good substitute.'
               (WS-U-T-B-0000000070.p.13.s.3)

     c.  *… als je   naar een oplossing zoekt die perfect   aansluit bij*
             if you to a    solution search that perfectly fits      to
         *je  bancaire behoeften.*
         your banking needs
         '…if you are searching a solution that fits your banking needs perfectly.'
               (WR-P-P-G-0000229626.p.13.s.1)

The production perspective proposed that *naar* presents a way to buy time for the producer to formulate a complex theme. When the theme precedes the verb, however, this purchase comes at a serious cost. Only a limited number of Dutch verbs frequently occur with a prepositional object with *naar*, at least compared to the number of verbs that appear with a direct object without *naar*, i.e. that occur in the transitive construction. Using *naar* would therefore force the producer to already think about which verb s/he is going to use. The planning scope of producers is limited, and the longer and more complex the upcoming theme argument, the less cognitive resources are available to simultaneously consider the choice of verb (Gleitman et al. 2007; Konopka 2012). Meanwhile, if the producer chooses to realize the theme as a bare noun phrase, s/he can postpone the choice of verb until after the theme is completed. Moreover, if the producer does already decide on the future verb while building the upcoming complex theme, s/he will be forced to retain this verb in working memory until s/he has completed the

formulation of the theme. Leaving this choice until later would allow him/her to free up this working memory.

An example with a complex preverbal theme is given in (71). When the producer includes *naar* in (71), his or her choice of verb will be limited to *zocht* 'searched' and perhaps *streefde* 'strove'. In other words, s/he would have to consider the choice of verb, right when facing the arduous task of planning the complex theme. Meanwhile, is the producer does not include *naar*, the choice of verb can be left for the future. In (71), reasonable options to finish the sentence would include *zocht* 'searched', but also *wilde volgen* 'wanted to follow', *probeerde te vinden* 'tried to find', *nastreefde* 'pursue', etc.

71. *De Wereldraad   van Kerken   heeft dat niet gedaan, omdat hij*
    The World_Council of   Churches has  that not  done,    because he
    *van begin af aan (naar) een derde weg tussen  het communistische*
    of  start off  on  (to)  a    third  way between the communist
    *oostblok     en   het vrije kapitalistische westen  zocht.*
    Eastern_bloc and  the free capitalist      West    searched.
    'The World Council of Churches has not done that, because, from the very beginning, it was searching for a third way between the communist Eastern bloc and the free, capitalist West.'
    (WR-P-P-G-0000103341.p.3.s.3)

To sum up, when the theme precedes the verb, more complex themes are likely to elicit the use of the variant without *naar*, viz. the transitive variant. Conversely, in instances where the theme is not complex, the producer is hardly under any processing pressure, and s/he might very well contemplate the choice of verb early on and choose to include *naar*. Meanwhile, when the verb precedes the theme, the verb is evidently already decided on at the moment when the choice of variant needs to be made. Therefore, producing *naar* does not put any extra burden on the producer's processing facilities, and can still be used to buy time to produce a complex theme without any extra cost associated with it. We therefore predict that there should be a negative correlation between THEME COMPLEXITY and the likelihood of *naar* when the theme precedes the verbs, and a positive correlation when the verb precedes the theme.

Taking the channel perspective, a parallel reasoning can be made. In cases where the theme precedes the verb, the presence of the preposition *naar* limits the number of verbs that may follow. Hence *naar* makes the following verb more predictable and therefore reduces its information content. Of course, this information does not just disappear: it is rather transferred over from the verb to

the preposition. In other words, the preposition literally signals a lot of information about the verb that is to follow.

This means that in instances where the preposition precedes the verb, the preposition already carries a lot of information. Combining such an informationally heavy preposition with a complex, informationally heavy theme would lead to a peak in information density, which should be avoided. Instead, combining the heavy preposition with a simple, informationally light theme would smooth out the information density. As such, we should expect *naar* 'to' to be preferred in front of simple themes, and be omitted in front of complex themes.

Of course, this reasoning only holds for instances where the theme precedes the verb. When the verb precedes the theme, the preposition evidently cannot signal any information about the verb, because the verb is already known at that point. As such, the now informationally light preposition can nicely combine with complex, informationally heavy themes. This leads to the following prediction, which is identical to the prediction made by the production hypothesis. There should be a negative correlation between THEME COMPLEXITY and the likelihood of *naar* when the theme precedes the verbs, and a positive correlation when the verb precedes the theme.

Finally, the comprehension perspective stated that *naar* functions as a signpost for the comprehender that simplifies the parsing of a complex theme. Such a signpost would be especially useful if a complex theme precedes the main verb, since in that case, it already gives considerable information about the verb that is to follow, as argued above. Because the main verb for a large part determines the structure of the entire sentence, knowledge of this verb would further simplify parsing to a great extent (Müller 2006; Müller and Wechsler 2014).

Put differently, *naar* 'to' would be extra effective in reducing the parsing effort of the comprehender when the theme precedes the verb. When the verb precedes the theme, it would still be helpful, but comparatively less crucial. As such, we formulate the following hypothesis. There should be a strong positive correlation between THEME COMPLEXITY and the likelihood of *naar* when the theme precedes the verbs, and a weaker positive correlation when the verb precedes the theme.

To sum up, our three perspectives make different predictions on how the correlation between complexity and explicitness behaves in different linguistic contexts. In particular, we have argued that the relevant distinction will be one between a context where the theme precedes the verb and one where the verb precedes the theme.

All three perspectives relate to mechanisms that are very general in nature and whose influence should not be dependent on the choice of preposition, verb or object. In other words, the complexity hypotheses should operate at the highest level of abstraction, and their predictions are not restricted to any particular verb or preposition.

This has two consequences. First, the complexity hypotheses can in principle be tested anywhere in our alternation, provided that we can properly keep lectal and semantic factors under control. Second, controlling for the influence of complexity while testing the lectal and semantic hypotheses is straightforward: we merely need to add to the regression models the variables THEME COMPLEXITY and VERB-THEME ORDER, i.e. a variable that distinguishes between instances where the verb precedes the theme and those where the theme precedes the verb, as well as an interaction between both variables. Conversely, controlling for the lectal and semantic factors while testing the complexity hypotheses is harder, since we do not yet know at which level of abstraction these factors operate.

As such, we will first test the lectal and semantic hypotheses in Chapter 6, and then try to track down any remaining meaning differences by applying data-driven analyses in Chapter 7. Only then, in Chapter 8, will we decide where and how we have the best chance of properly keeping the lectal and semantic factors under control in order to get a clear sight on the effects of the complexity hypotheses. Our decision to first test the lectal and semantic hypotheses and only then turn to the complexity hypotheses is thus merely a practical choice: it does not imply any sort of relative ranking of importance or hierarchy in these hypotheses.

### 4.3.2.3    Suitability of the data

The question could be raised whether a dataset drawn from a corpus of written language is suitable to test hypotheses relating to the Complexity Principle. This subsection argues that it is. The Complexity Principle has been observed both in spoken and written language (a.o. Bouma, 2017; Rohdenburg, 1996, 2016; Shank et al., 2016b). In the present study, we are first and foremost looking to explain why the Complexity Principle holds in written language. Still, we currently see no compelling reasons to assume that there are fundamentally different explanations for the Principle in spoken versus written language. In fact, there is ample research showing that findings based on written language are generally in accordance with findings from spoken language with regard to the relation between explicitness and complexity (Grondelaers 2000; Grondelaers, Speelman and Geeraerts 2003; Jaeger and Wasow 2005; Jaeger et al. 2005).

Still, even if future research would reveal fundamental differences, it is not the case that written language processing is a priori less interesting than spoken language processing; this would simply limit the relevance of our research to the former. In using written data to study the influence of complexity, we follow earlier studies that include Bloem, Versloot, & Weerman (2017), Gennari & Macdonald (2009), Gries (2002), Grondelaers et al. (2009), Jaeger (2011), MacDonald & Thornton (2009), Rohdenburg, (2016), Roland et al. (2006) and Willems & De Sutter (2015).

What does this choice for written data mean for the three perspectives introduced in the previous subsections? In general, this choice is favorable for the comprehension perspective. First, it benefits the comprehension perspective in that we expect writers to bear in mind the ease with which their readers read their texts, at least to a greater extent or more explicitly than speakers would take into consideration the comprehension processing required from their hearers. As such, written language would be more prone to tendencies that reduce effort in comprehension processing.

Second, the choice for written data is harmful to the production and channel perspectives. Regarding the production perspective, its reasoning primarily relates to spoken language. When producing spoken language, the producer is typically under time constraints. In dialogues or multilogues, failure to produce utterances in due time may result in another participant taking their turn to speak. In monologues, such failure may result in losing the attention of the audience. In written language, time constraints are typically less pressing – if they are present at all. As such, the production-driven explanation needs the extra assumption that the (probabilistic) grammar of language users is first and foremost shaped by their experiences in spoken language, since this generally forms the majority of their linguistic input, and that this same grammar is then employed when processing written language. The correlation between complexity and explicitness in written language would then be a second order effect, i.e. an effect that is retained even when its original cause is not directly present, because it has become entrenched in probabilistic grammar. Such second-order effects have also been demonstrated in morphology (cf. Pijpops and Van de Velde's 2016 analyses of the Dutch partitive genitive).

Regarding the channel perspective, the information channel of written language is probably less prone to noise than the channel of spoken language. Therefore, it would arguably be associated with a higher optimal degree of information density.[53] As such, the channel would generate less pressure to use optional markers in complex environments in the case of written language than in the case of spoken language. Still, the channel of written language would still be associated with some optimal degree of information density. As such, the reasoning behind this perspective still holds. To sum up, the choice for written data results in a conservative research design regarding the production and channel perspective.

---

[53] We say *arguably*, because spoken language allows for much more multimodal ways of reinforcing the signal, e.g. through the use gestures and facial expressions, which could in principle result in a higher optimal degree of information density in spite of there being more noise (Cover and Thomas 1991). In addition, if information is lost in the channel of written language, the comprehender is often not in the position to signal this loss to the producer, while in the case of spoken language, this is typically still possible. Information loss in written language would hence be more definitive.

Still, to make our dataset suitable to testing the complexity hypotheses, it is crucial to exclude a particular subset of data from our analyses, viz. those instances where the theme is placed in postfield position. To explain this, a short introduction into Dutch sentence structure is required. Dutch sentence structure functions a lot like its German counterpart, and is also characterized by a bipolar structure (i.e. the so-called *Klammernstruktur*, see Haeseryn et al. 1997: 1225–1400; Zifonun, Hoffmann and Strecker 1997: 1498; König and Gast 2018: 188–214; Zwart 2011: 25–79).

This structure is laid out in Table 12. Bare noun phrases such as subjects or direct objects are grammatically limited to the prefield before the first verbal pole (72a) or the midfield between the poles (72c). They cannot grammatically be placed in the postfield, i.e. the position behind the second verbal pole (72e), the only exception being when they are realized as a subordinate clause. By contrast, prepositional phrases such as the prepositional object have access to the prefield, midfield and postfield (72b,d,f). This means that if the language user really wants to place a nominal theme argument in postfield position, employing the prepositional variant is the only way to achieve this.

|     |    | Prefield | 1st verbal pole | Midfield | 2nd verbal pole | Postfield |
|-----|----|----------|-----------------|----------|-----------------|-----------|
| 72. | a. | *Adem* breath | *heb* have | *ik* I | *gehapt* gasped | ∅ |
|     | b. | *Naar adem* | *heb* | *ik* | *gehapt* | ∅ |
|     | c. | *Ik* | *heb* | *adem* | *gehapt* | ∅ |
|     | d. | *Ik* | *heb* | *naar adem* | *gehapt* | ∅ |
|     | e. | *\*Ik* | *heb* | ∅ | *gehapt* | *adem* |
|     | f. | *Ik* | *heb* | ∅ | *gehapt* | *naar adem* |

'I gasped for breath.'

Table 12: Placement options for the theme participant in the Dutch sentence structure. Nominal constituents cannot be placed in the postfield.

When the theme argument is complex, the pressure to place it in postfield would mount (cf. the Principle of End Weight, Behaghel 1909; Wasow 2002), which would in turn increase the usage of the prepositional variant. It is in principle possible

that this is the only reason for an observed correlation between theme complexity and the probability of the prepositional variant – which would actually be at odds with all three viewpoints. If the reasoning behind any of the three viewpoints is correct, however, then their predictions should persist even when the postfield instances are excluded from the dataset. As such, we need to exclude these instances when testing the complexity hypotheses in Chapter 8.

## 4.4 Overview of the predictions

We end this chapter by providing an overview of all hypotheses and predictions that were formulated here. The hypotheses are named *Lec1, Lec2, Sem1,...* where *Lec* stands for lectal, *Sem* for semantic and *Com* for complexity. For each semantic hypothesis, several predictions were made. These are further coded as 1, 2a and 2b, where 1 indicates that the prediction pertains to the same level of abstraction as where the hypothesis is formulated, and 2 indicates that it pertains to the next underlying level of abstraction.

Note that several of these hypotheses make conflicting predictions. The predictions of the complexity hypotheses conflict most starkly, with on the one hand Com1 and Com2 and on the other Com3. The same holds – albeit less starkly – for the semantic hypotheses, with on the one hand Sem1 and on the other Sem2, Sem3 and Sem4. Sem1 predicts that the alternation between the transitive and the *naar*-construction expresses the same meaning difference, irrespective of the verb. In contract, Sem2, Sem3 and Sem4 predict that the expressed meaning differences are dependent upon the verb. As such, we fully expect some of these hypotheses to be refuted. The question is only which ones. The next chapter presents the manual checking of the data that will be used to test these hypotheses.

**Lec1**:  The predictive quality of the models fitted on the Netherlandic data will generally be higher than those fitted on the Belgian data.

**Lec2**:  The predictors relating to lexical biases will cause a greater increase in predictive quality in the Netherlandic models than in the Belgian models.

**Sem1**: The *naar*-construction expresses directionality, the transitive construction does not.

    1.    The measure VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION will be positively correlated with the probability of the prepositional variant in the Netherlands and Belgium.

    2.    The measure OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION will be positively correlated with the probability of the prepositional variant in the Netherlands and Belgium, among instances of a single verb.

**Sem2**: The meaning of 'desire' has specialized to 'demand' for the transitive *verlang*-construction, and to 'long for' for the *verlang-naar*-construction.

    1.    The measure SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION will be positively correlated with the probability of the prepositional variant in the Netherlands and Belgium, among the instances of the verb *verlangen* 'desire'.

    2a.   In the Netherlands and Belgium, among instances of the verb *verlangen* 'desire' and the theme root *ding* 'thing', those instances that involve longing for a thing will prefer the prepositional variant, whereas those demanding a thing will prefer the transitive variant.

    2b.   In the Netherlands and Belgium, among instances of the verb *verlangen* 'desire' and the theme root *tijd* 'time', those instances where *tijd* 'time' acts as a count noun will prefer the prepositional variant, whereas those where *tijd* 'time' is a mass noun will prefer the transitive variant.

**Sem3**: The meaning of 'gauge' has specialized to 'directly judging' for the transitive *peil*-construction, and to 'gauging by asking' for the prepositional *peil-naar*-construction in Belgium.

    1.    The measure SEMANTIC COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION will be positively correlated with the probability of the prepositional variant in Belgium, among the instances of the verb *peilen* 'gauge'.

    2a.   Among Belgian instances of the verb *peil* 'gauge' and the theme root *stemming* 'mood', those that involve asking about the mood will prefer the prepositional variant, whereas those where the mood is directly judged will prefer the transitive variant.

2b. Among Belgian instances of the verb *peilen* 'gauge' and the theme root *reactie* 'reaction', those that involve asking about a reaction will prefer the prepositional variant, whereas those where a reaction is directly judged will prefer the transitive variant.

**Sem4**: The meaning of the transitive *zoek*-construction inclines more towards 'seeking to make or to acquire something', whereas the meaning of the *zoek-naar*-construction inclines more towards an act of literally looking for something.

1. The measure SEMANTIC COHERENCE TO THE *ZOEK-NAAR*-CONSTRUCTION will be positively correlated with the probability of the prepositional variant in the Netherlands and Belgium, among the instances of the verb *zoeken* 'search'.

2a. In the Netherlands and Belgium, among instances of the verb *zoeken* 'search' and the theme root *slachtoffer* 'victim', those instances where the agent is a helper will prefer the prepositional variant, whereas those where the agent is an aggressor will prefer the transitive variant.

2b. In the Netherlands and Belgium, among instances of the verb *zoeken* 'search' and the theme root *woord* 'word', those instances where the agent is searching for specific words will prefer the prepositional variant, whereas those where agent is searching for non-specific words, e.g. to try to express a proposition, will prefer the transitive variant.

**Com1**: The Complexity Principle is primarily caused by constraints in language **production**. Therefore, there should be a negative correlation between THEME COMPLEXITY and the likelihood of *naar* 'to' when the theme precedes the verbs, and a positive correlation when the verb precedes the theme.

**Com2**: The Complexity Principle is primarily caused by constraints in the language **channel**. Therefore, there should be a negative correlation between THEME COMPLEXITY and the likelihood of *naar* 'to' when the theme precedes the verbs, and a positive correlation when the verb precedes the theme.

**Com3**: The Complexity Principle is primarily caused by constraints in language **comprehension**. Therefore, there should be a strong positive correlation between THEME COMPLEXITY and the likelihood of *naar* 'to' when the theme precedes the verbs, and a weaker positive correlation when the verb precedes the theme.

# 5    Data selection

In the Chapter 3, we have selected the verbs that alternate between the transitive and prepositional intransitive construction, extracted their instances and decided to focus on those verbs alternating with *naar* 'to'. We now need to decide which of these instances to exclude from our dataset. This is not an easy task, since the alternation does not take place in a vacuum, but forms part of a larger language system. In our case, the transitive and prepositional constructions interact with several other constructions, and these interactions need to be considered when deciding which instances to retain in the dataset. We begin this section by outlining some general guidelines and considerations regarding the distinction between the instances to be removed and the instances to be retained, and then list the types instances that were removed, as well as those that were retained after some doubt and careful deliberation.

## 5.1    Categorical contexts, alternation factors and knock-on effects

To decide which instances to exclude, we need to distinguish between categorical contexts, alternation factors and knock-on effects. Categorical contexts relate to phenomena that directly causally determine the use of the variants, i.e. that categorically preclude the use of one of the variants. Such instances should be removed from the dataset under the definition of alternations as an individual's choice points (see Section 2.1). An example would be instances where one of our alternating verbs is coordinated with another verb with which it shares its theme argument, as in (73). The coordination in (73) precludes the use of the prepositional variant, since the verb *verdienen* 'deserve' cannot be used with a prepositional object with *naar* 'to'. Hence, such instances would have to be removed.

By contrast, alternation factors are phenomena that may causally influence though not completely fix the use of the variants. The instances where these factors are at work should be retained in the data and the factors may be used to predict the choice of variant in the analysis. An example would be the complexity of the theme argument (see Section 4.3).

73. *Als je eenmaal bewust een gedachte kiest (…) kun je nu*
    When you once consciously a thought choose can you now
    *op weg gaan naar de vrijheid die je verlangt en verdient.*
    on road go to the freedom that you desire and deserve
    'Once you consciously choose a thought (…), you can now follow your path
    to the freedom that you desire and deserve.'
    (WR-P-P-B-0000000020.p.865.s.4)

Finally, knock-on effects are phenomena that are causally determined by the choice
of variant. Instances where these can be at play should be retained in the analysis,
but the researcher may of course not employ these knock-on effects to predict the
alternation, as such a prediction would be circular. An example would be the use
of the pronominal adverb in (74a). The choice between a simple combination of the
preposition and the pronoun, e.g. *naar het* 'to it', and a pronominal adverb, e.g.
*ernaar* 'thereto', is only available when the prepositional object is employed (see
Haeseryn et al. 1997: 490–503). This choice involves the forms *naar het* 'to it' –
*ernaar* 'thereto', *naar dat/die* 'to that' – *daarnaar* 'thereto', *naar dit/deze* 'to this' –
*hiernaar* 'hereto', *naar wat* 'to what' – *waarnaar* 'whereto', *naar iets* 'to something'
– *ergens naar* 'somewhere to', *naar alles* 'to everything' – *overal naar* 'everywhere
to', and *naar niets* 'to nothing' – *nergens naar* 'nowhere to'. In our analyses, we
will never distinguish between a realization as preposition and pronoun, as in (74b)
and pronominal adverb, as in (74c). The alternation between preposition plus
pronoun and pronominal adverb falls beyond the scope of the present study and
may be subjected to an alternation study in its own right.

74. a. *Ik weet dat jij ernaar verlangt.*
       I know that you thereto desire
       'I know that you desire it.'
       (WR-P-E-G-0000007957.p.318.s.1)

    b. *U verlangt naar iets, maar krijgt het niet.*
       you desire to something but get it not
       'You desire something, but don't get it.'
       (WR-P-P-B-0000000479.p.152.s.1)

    c. *Verlangt u nog ergens naar?*
       desire you still somewhere to
       'Do you desire anything else?'
       (WR-P-P-G-0000173987.p.43.s.2)

This three-way distinction between categorical contexts, alternation factors and knock-on effects may sound straightforward, but it is actually often difficult to apply in both theory and praxis. We already discussed some of the difficulties in distinguishing between categorical contexts and alternation factors in Section 2.1. In order to theoretically distinguish knock-on effects from alternation factors and categorical contexts, we need to assume some order in which linguistic choices are made.

For instance, in order for the complexity of the theme argument to have any bearing on the choice of variant, this complexity needs to exist or at least needs to be approximated in some way, before the choice of variant is made (cf. Van Beveren, Colleman and De Sutter 2019: 214). If the language producer first decides on the choice of variant, and only thereafter begins to plan the theme argument, then the complexity of this argument of course cannot affect the choice of variant. By naming THEME COMPLEXITY as a (potential) alternation factor, we therefore also assume that some approximation of its value already exists when the choice of variant is made.

Conversely, the choice between a combination of a preposition and a pronoun, e.g. *naar dit* 'to this' and a pronominal adverb, e.g. *hiernaar* 'hereto', is only available if the prepositional variant is chosen. Hence, the choice of variant needs to precede the choice between *naar dit* 'to this' and *hiernaar* 'hereto'. This is why we have called this choice a knock-on effect. However, we do not yet know the exact order of all linguistic choices in the current state of linguistics (for research on this topic, see Rueschemeyer & Gaskell 2018: 291−545). As a result, some of the decisions that are detailed in the remainder of this chapter, can certainly be contested.

Let us immediately illustrate this by revisiting the example of coordination. It is possible to rephrase (73) to employ the prepositional variant, provided that theme coordination is relinquished, as in (75a). Now, it would be conceivable that if all probabilistic factors line up to promote the use of the prepositional variant, this could indeed lead the language user to relinquish theme coordination and express his/her sentence as (75a). Still, it was decided to consider the relation between theme coordination and the alternation as a categorical context, rather than an alternation factor or a knock-on effect, and hence, the 732 instances such as (73) were removed. Note that an instance like (75a) would not have been removed, because an alternative as in (75b) would of course be possible. Rare instances where both coordinated verbs actually allowed the alternation, such as (75c), were also removed. This was done for consistency.

75. a. *Als je eenmaal bewust een gedachte kiest (…) kun je*
    when you once consciously a thought chooses can you
    *nu op weg gaan naar de vrijheid waarnaar je verlangt en*
    now on road go to the freedom whereto you desire and
    *die je verdient.*
    that you deserve
    'Once you consciously choose a thought (…), you can now follow your
    path to the freedom that you desire and that you deserve.'

    b. *Als je eenmaal bewust een gedachte kiest (…) kun je*
    When you once consciously a thought chooses can you
    nu op weg gaan naar de vrijheid die je verlangt en die
    now on road go to the freedom that you desire and that
    je verdient.
    you deserve
    'Once you consciously choose a thought (…), you can now follow your
    path to the freedom that you desire and that you deserve.'

    c. *Maar die kunnen niet bieden wat de gemiddelde alleenstaande*
    but they can not offer what the average single
    *zoekt en verlangt.*
    searches and desires.
    'But they cannot offer what the average single searches and desires.'
    (WR-P-P-G-0000171046.p.3.s.2)

Five more phenomena need to be discussed in this regard: the interaction with
constructions of negation, word order, the interaction with the passive, semantic
nuances and lexical biases.

The choice of construction for negation, viz. implicit vs. explicit, was considered
a knock-on effect (see Haeseryn et al. 1997: 1639–1659). The use of implicit
negation is more natural for transitive occurrences with indefinite themes, as in
(76a), while the reverse holds for prepositional occurrences in comparable contexts,
as in (76b). Still, both variants can employ both types of negation under certain
conditions, e.g. in a contrastive context, as in (77).

76. a. *Ik ben kwaad op mezelf dat ik die nacht (voor de finale) geen*
    I am angry on myself that I that night (for the finals) no
    *dokter heb gebeld.*
    doctor have called.
    'I am angry with myself for not calling a doctor that night, before the
    finals.'
    (WR-P-P-G-0000572981.p.2.s.2)

b.   *Ik verlang echt   niet naar een  goednieuwsdocumentaire over*
     I  desire   really not to  a    good_news_documentary about
     *lammetjes.*
     little_lambs
     'I do not desire a good news documentary about little lambs.'
              (WR-P-P-H-0000167553.p.1.s.6)

77.  *a.   Ik zocht     niet iemand    die goed is met naald en   draad*
          I  searched not  someone   that good is with needle and  thread
          *maar een grote  groep kan  inspireren (...)*
          but  a   big    group can  inspire
          'I'm not looking for someone who is capable with needle and thread,
          but someone who can inspire a large crowd...'
              (WR-P-P-G-0000402083.p.10.s.4)

b.   *Maar ik zocht    naar geen verontschuldigingen.*
     but  I  searched to   no   apologies
              (WR-P-P-G-0000656198.p.3.s.4)


Regarding word order, the placement options for nominal and prepositional constituents in the Dutch two-pole sentence structure are not identical, as explained in Table 12 in Subsection 4.3.2.3. The transitive construction cannot be used if the theme is placed in postfield position, except if the theme takes the form of a subordinate clause. In fact, the placement of prepositional constituents in the midfield vs. the postfield has been studied as an alternation in its own right by Willems & De Sutter (2015).

There are three logical possibilities: either the choice between midfield placement and postfield placement of the theme is made before the choice between the transitive and prepositional variant, or after it, or both choices are made simultaneously. In the first case, if the postfield is chosen, it is no longer possible to use the transitive variant. Hence the postfield should be considered a categorical context, and instances where the theme is placed in the postfield should be removed from the dataset.

In the second case, the placement of the theme in midfield or postfield cannot have any bearing on the choice between transitive and prepositional variant, since it has not been determined yet. The possibility of placing the theme in postfield position would then be considered a mere knock-on effect of the choice between transitive and prepositional construction, just like the choice between *naar dit* 'to this' and *hiernaar* 'hereto'. As such, the instances in the postfield would be regarded as genuine instances where the prepositional variant was explicitly chosen over the transitive variant and these instances would need to be retained in the dataset.

In the third case, both alternations should be collapsed into one super-alternation with levels like *transitive-midfield, prepositional-midfield* and *prepositional-postfield*, and *even prepositional-prefield*, etc. We have chosen to go with the first possibility, and hence excluded all postfield instances such as (78) from the dataset before the manual checking. This was done by excluding all instances of the prepositional variant where the verb was an infinitive or a participle, or a finite verb in a verb-final clause, and preceded the beginning of the theme.[54] Instances where the theme constituent begins before the verb but ends behind it, as in (79), where not excluded. Note that even if we would have chosen to consider postfield placement a knock-on effect, we would still need to exclude the postfield instances while testing the Complexity Hypotheses, as this is dictated by their predictions (see Subsection 4.3.2.3), and we would need to exclude them from manual annotation, as we cannot easily blind them for the choice of variant (see Sections 6.3 and 7.3).

78. *Wij gaan vissen naar garnalen.*
    we  go   fish   to    shrimps
    'We're going fishing for schrimps.'
                (WR-P-E-G-0000004444.p.363.s.1)

79. *Waarop    mijn moeder onmiddellijk naar een  oom   telefoneerde die*
    whereon   my   mother immediately  to   an   uncle phoned       that
    *burgemeester was van  ons  dorp.*
    mayor        was of   our  town
    'Whereupon my mother immediately called an uncle who was the mayor of our town.'
                (WR-P-P-G-0000203589.p.14.s.14)

In the midfield, nominal and prepositional complements still have slightly different positional preferences, with prepositional complements tending more towards the end of the midfield than nominal complements (Haeseryn et al. 1997: 1306–1354). This generally does not lead to a difference in practice, as in (80). Sometimes, however, it does. For instance, prepositional complements prefer to be placed behind prepositional adjuncts in the midfield, while nominal complements do not. As such, the direct object can be placed fairly naturally in front of *in de eerste plaats*

---

[54] By 'verb-final clause', we mean all clauses with the CAT-tag *whrel, cp, rel, ssub, whsub, oti, ti, ehi, svan*, or *inf* (van Noord 2006; van Noord et al. 2013). Instances where the finite verb preceded the beginning of the theme in clauses with the CAT-tag *whq* were retained at this point, and subjected to manual checking, as these did not reliably identify verb-final clauses.

'in the first place' in (81a), while for the prepositional object, this position feels more marked, as in (81b) (cf. Subsection 2.2.2). We consider this to be a knock on effect, and retain instances like (81a-b) in our dataset.

We have two reasons for considering the placement options within the midfield a knock-on effect. First, the placement difference within the midfield is much less dramatic than between the midfield and the postfield, and we believe that it can be easily overwritten by the influence of an alternation factor. That is, if one or several important alternation factors would point towards the prepositional variant, we suspect that this could easily cause the language user to produce (81b) or (81d) instead of (81a). Second, this placement difference is a tendency rather than a categorical difference. The direct object can well be placed behind a prepositional adjunct, as in (81c) or (82a), or the prepositional object in front of a prepositional adjunct as in (81b) or (82b).

80.  *Dus toen   we (naar) popmuziek  voor deze Klinkklaar  gingen zoeken,*
     so   when  we (to)    pop_music  for  this Klinkklaar  went    search
     *kregen we    een probleem.*
     got     we    a   problem
     'So when we went searching for pop music for this Klinkklaar, we
     encountered a problem.'
                    (WS-U-T-B-0000000706.p.1.s.2)

81.  a.  *Computergebruikers zullen het  eurosymbool in de     eerste*
         computer_users      will   the  euro_symbol in the    first
         *plaats zoeken  op hun toetsenbord.*
         place  search   on their keyboard
         'The users of computers will in the first place search for the euro-
         symbol on their keyboard.'
                    (WR-P-P-H-0000050539.p.4.s.1)

     b.  *Computergebruikers zullen naar het eurosymbool in de eerste*
         computer_users      will   to   the euro_symbol in the first
         *plaats zoeken   op hun  toetsenbord.*
         place  search    on their keyboard
         'The users of computers will in the first place search for the euro-
         symbol on their keyboard.'
                    (feels marked)

    c.   *Computergebruikers zullen in de eerste plaats het eurosymbool*
        computer_users      will   in the first   place  the  euro_symbol
        *zoeken op hun toetsenbord.*
        search   on their keyboard
        'The users of computers will in the first place search for the euro-symbol on their keyboard.'

    d.   *Computergebruikers zullen in de eerste plaats naar het*
        computer_users      will   in the first   place  to     the
        *eurosymbool zoeken op hun toetsenbord.*
        euro_symbol search   on their keyboard
        'The users of computers will in the first place search for the euro-symbol on their keyboard.'

82.  a.   *Mexx en Esprit, niet direkt kleine*
        Mexx and Esprit not immediatelysmall
        *konfektiebedrijven, zoeken samen met*
        ready-made_clothes_companies search together with
        *ontwikkelingsorganizaties toeleveranciers in de Derde Wereld*
        development_organizations suppliers      in the Third World
        *die relatief goede werkomstandigheden in acht nemen.*
        that relatively good working_conditions  in condition take
        'Mexx and Esprit, not immediately small clothing companies, are working together with development organizations to search suppliers that maintain relatively good working conditions.'
                (WR-P-P-H-0000000797.p.19.s.2)

    b.   *In het dorpje Kato Achaia, (…) zoeken* reddingsploegen naar   een
        in the village Kato Achaia      search  rescue_teams    to     a
        *meisje van negen in de resten van een huis.*
        girl   of   nine  in the remains of   a     house
        'In the village of Kato Achaia, rescue teams are searching for a nine year old girl in the remains of a house.'
                (WS-U-E-A-0000113660.p.1.s.1)

We now turn to passive instances, as in (83). The passive has been dealt with in a variety of ways in theories of language. A first possibility, compatible with the lexical approaches to argument realization, is to employ lexical rules that take verbs with an active transitive or prepositional intransitive argument structure and license passive uses of those verbs (cf. Sag 2012: 115–116; Müller & Wechsler 2014). Following this analysis, instances such as (83) should be retained in the dataset, as

they represent genuine uses of the transitive (83a-c) and prepositional (83d-f) variants of the verbs.

83. a. *Toen de buren     er    zondagmorgen   genoeg van hadden,*
       when  the neighbors there Sunday_morning enough of    had
       *werd   de politie gebeld.*
       was    the police  called
       'When the neighbors had had enough on Sunday morning, the police
       was called.'
                  (WR-P-P-G-0000531035.p.2.s.2)

    b. *Ik weet  ook dat als    het slecht loopt, er    altijd excuses*
       I  know  also that when it  bad    runs   there always excuses
       *gezocht    worden   om de dip te  verklaren.*
       searched   are         to the dip to  explain
       'I know that when things do bad, there are always excuses being
       sought to explain the dip.'
                  (WR-P-P-H-0000134810.p.33.s.5)

    c. *Ze    worden gezocht met speurhonden.*
       they are       search  with tracker_dogs
       'They are being searched with tracker dogs.'
                  (WS-U-E-A-0000212722.p.1.s.5)

    d. *We merken immers dat  lang niet altijd    naar de politie wordt*
       we notice  after_all that long not  always to    the police is
       *gebeld als     ergens      een alarm afgaat.*
       called when   somewhere an    alarm goes_off
       'After all, we notice that most of the time, the police isn't called when
       an alarm goes off somewhere.'
                  (WR-P-P-G-0000663925.p.4.s.3)

    e. *Naar excuses werd er     niet gezocht   door trainer  Bloemen.*
       to    excuses was there  not  searched  by    trainer  Bloemen
       'Excuses were not sought after by trainer Bloemen.'
                  (WR-P-P-G-0000531035.p.2.s.2)

    f. *Naar hen    wordt nog gezocht.*
       to    them  is     still searched
       'They are still sought after.'
                  (WS-U-E-A-0000000547.p.6.s.30)

A second possibility is to posit a separate passive construction that may freely combine with the transitive construction or the *naar*-construction (cf. Michaelis 2006). For instance, in order to form the full utterance in (83a), the language producer would combine the transitive construction with the passive construction, just like he/she combines it with a large number of other constructions, such as the syntactic NP-construction, the lexical *buur*-construction, the plural *en*-construction, etc. Meanwhile in (83c), the *naar*-construction would be combined in the same way with the passive construction – and a whole number of other constructions – to form the full utterance.

Following this proposal, we would also need to retain instances such as those in (83) in the dataset. The fact that these utterances happen to contain a passive construction while other instances do not, would be no more of a valid reason to exclude them than that they happen to contain a lexical *altijd*-construction (83b,d), a morphological compound-construction (83a, *zondagmorgen*), an explicit negation construction (83d,e), etc., while other instances do not.

A third possibility is to systematically distinguish between an active transitive construction and a passive transitive construction, as well as an active *naar*-construction and passive *naar*-construction, etc. (cf. Goldberg 1995: 56–57, 138–139; van Trijp 2011). The active and passive variants of each construction can then be related to one another through an overarching transitive construction respectively *naar*-construction (thereby forming allostructions, cf. Cappelle 2006), or through strong, productive links between both constructions (Goldberg 1995: 138–139).

Following this proposal, it could both be argued to exclude or retain instances like (83) in the dataset. On the one hand, one could claim to be solely interested in the alternation between the active transitive and the active *naar*-construction, and hence exclude instances like (83) as they do not instantiate either construction. On the other hand, one could argue to be interested either in the alternation between the overarching transitive and *naar*-construction, or in the alternation between, on one side, the closely interlinked active and passive transitive constructions and on the other side, the closely interlinked active and passive *naar*-construction.

We conclude that there are at least more – if not better – arguments to retain the passive instances than to exclude them, and that is what we chose to do. Of course, in passive uses of the transitive variant, the theme argument takes subject function and the verb has to formally agree with it, which is not the case for the prepositional variant. We consider this difference in verbal agreement, e.g. between (83b) and (83e), and the difference between the subject and object form of the pronoun, as in (83c) and (83f), to be knock-on effects of the variant choice.

We take an explicit interest in meaning differences between the variants, and want to test a number of semantic hypotheses. As such, we cannot a priori ban all meaning differences from our dataset. Only in cases where there is one lexical

sense that can be delineated more or less cleanly from the other instances and that strictly allows one variant, will these instances be excluded from the dataset.

Finally, as explained in Section 1.2, we treat lexical biases as an explanandum rather than an explanans. As such, a lexical bias as such is not considered a sufficient reason for exclusion from the dataset. Moreover, it appears that the lexical biases in our dataset are rarely if ever strictly categorical and can be overcome if the lexeme is present in sufficient numbers, or perhaps if some alternation factor exerts sufficient influence. For instance, the object *aansluiting* 'connection' seems to occur most often with the transitive variant, as in (84a), but does occur with the prepositional variant as well, if only rarely, as in (84b).

84. a.   *De geschiedenis van het bedrijf,   dat  met  zijn functionele*
         the history        of  the company that with its    functional
         *en   sobere ontwerpen altijd aansluiting zocht     bij   het*
         and sober  designs       always connection searched with the
         *Modernisme, loopt  als een  rode  draad door    de tentoonstelling.*
         Modernism  runs   as  a    red  thread through the exhibition
         'The history of the company, that with its functional and sober designs always sought to connect with Modernism, forms a reoccurring theme throughout the exhibition.'
                   (WR-P-P-G-0000070837.p.27.s.2)

    b.   *Eigenlijk zochten    de christenen altijd  wel  naar aansluiting*
         actually   searched  the Christians always part to    connection
         *met  de socialisten.*
         with the socialists
         'Actually, the Christians always sought to connect with the socialists.'
                   (WR-P-P-G-0000636138.p.4.s.2)

## 5.2   Results of the manual data selection

The current section lists all choices that were made during the manual selection of the data, which took into account the considerations and guidelines laid out in the previous section and in Section 2.1. We first list all types of instances that were excluded for all verbs, followed by those types of instances that were retained in the dataset after some doubt. Next, we do the same for the various verbs individually. In addition to data selection, we also corrected the values for our variables of interest. That is, genuine instances where the theme argument was wrongly delineated, the wrong theme root was selected, or that were wrongly

labeled regarding the relative position of theme and verb or the choice of variant were corrected and retained in the dataset. In the case of true attachment ambiguities in the delineation of the theme, the delineation assigned by the Alpino-parser was retained. An example is (85), where *in de nieuwe wet* 'in the new law' can be attached both to the noun *mazen* 'loopholes' and the verb *zoeken* 'search'. Here, Alpino had chosen to attach it to *mazen* 'loopholes', and this choice was retained.

In total, 117,697 instances were manually checked. These are all instances of the verbs alternating with the preposition *naar* 'to', viz. *verlangen* 'desire', *peilen* 'gauge', *zoeken* 'search', *grijpen* 'grab', *graaien* 'grasp', *grabbelen* 'scramble', *happen* 'snap', *schoppen* 'kick', *bellen* 'phone', *telefoneren* 'phone', *opbellen* 'phone', *jagen* 'hunt' and *vissen* 'fish', where the theme argument is expressed and not placed in postfield position and for which the country of origin is known.

85. *De uitbaters van nachtwinkels zoeken nu al      naarstig naar*
    the managers of   night_shops   search   now already  diligently  to
    mazen   in  de nieuwe  wet.
    meshes  in  the new      law
    'Managers of night shops are already diligently searching for loopholes that may exist in the new law.' or
    'Managers of night shops are already diligently searching, in the new law, for loopholes.'
                          (WR-P-P-H-0000002059.p.6.s.1)

### 5.2.1     All verbs

#### 5.2.1.1     Removed instances

The following instances were manually removed for all verbs.

- Instances that were wrongly extracted or that were wrongly retained because of some parsing error, viz.:
    - 1276 instances where the verb was not the verb at issue or not even a verb at all, but that had been extracted because it had the wrong ROOT-tag or the wrong POS-tag.
    - 1079 instances where the theme argument was not expressed, but some other constituent had been wrongly identified as the theme.

- o 105 instances of a prepositional object that has a preposition different from *naar* 'to'.
  - o 252 instances where the prepositional object was placed in the postfield.
- 72 instances that were written in another language than Dutch. This sometimes happened for *verlangen* in German, for *happen* in English, or *bellen* in French.
- All 953 remaining 'double' instances. These are all instances that were clear cases of copying but that had not yet been removed automatically because of a minor difference in punctuation, word order, word choice etc., as in (86). Two instances of *zoeken* were actually identical except for the choice of variant, viz. (87). These were both retained. Of course, instances of sentences that can be expected to occur often in a similar form, such as in (88), were all retained.
- All 6680 resultative instances such as (89). It could be argued that these instances present prime examples of a conative meaning difference between the variants, and hence should be kept in the analysis (cf. Section 2.1). That is, the reason why these only appear in the transitive variant is because the transitive variant entails or implies a successful execution of the action expressed by the verb. Still, even if so, we would want to know whether this entailment or implication follows from the presence of the resultative constituent or from the use of the transitive variant. If the latter is the case, we should still be able to detect a conative meaning difference if all resultative instances are excluded.

86. a. *Nadat ze    de inhoud* **uit** *de kassa* **hadden genomen**, *zochten*
     after   they the content  out the till      had      taken    search
     *ze    tevergeefs naar een  kluis achteraan  in de winkel.*
     they vainly     to   a    safe at_the_back in the store
     'After taking the contents out of the till, they vainly searched for a safe in the back of the store.'
           (WR-P-P-G-0000212581.p.1.s.3)

    b. *Nadat ze    de inhoud* **van** *de kassa* **genomen hadden**, *zochten*
     after   they the content of   the till    taken    had      searched
     *ze    tevergeefs naar een  kluis achteraan  in de winkel.*
     they vainly     to   a    safe at_the_back in the store
     'After taking the contents of the till, they vainly searched for a safe in the back of the store.'
           (WR-P-P-G-0000212636.p.4.s.3)

c. *Het gerecht zoekt nu nog een meisje,dat in de nacht van*
the court searches now still a girl that in the night of
**Annicks verdwijning** *langs dezelfde route naar huis zou*
Annick's disappearance along the_same route to house would
*zijn gefietst.*
be biked
'Authorities are now still searching for a girl, that in the night of
Annick's disappearance would have biked home along the same route.'
(WS-U-E-A-0000239765.p.2.s.9)

d. *Het gerecht zoekt nu nog een meisje,dat in de nacht va*
the court searches now still a girl that in the night of
**de verdwijning van Annick** *langs dezelfde route naar huis*
the disappearance of Annick along the_same route to house
*zou zijn gefietst.*
would be biked
'Authorities are now still searching for a girl, that in the night of the
disappearance of Annick would have biked home along the same
route.'
(WS-U-E-A-0000239777.p.1.s.5)

87. a. *Het parket zoekt ook nog Lesley D, het liefje van*
the public_prosecutor searches also still Lesley D, the lover of
*Johri dat de helicopter inhuurde.*
Johri that the helicopter hired
'Authorities are also still searching for Lesley D., the lover of Johri,
who hired the helicopter
(WS-U-E-A-0000141104.p.1.s.7)

b. *Het parket zoekt ook nog naar Lesley D, het liefje*
the public_prosecutor searches also still to Lesley D, the lover
*van Johri dat de helicopter inhuurde.*
of Johri that hired the helicopter
'Authorities are also still searching for Lesley D., the lover of Johri,
who hired the helicopter
(WS-U-E-A-0000141166.p.1.s.8)

88. a. *Ik zal zien wat ik kan doen. Zoek liever de moordenaar.*
I will see what I can do search rather the killer
'I'll see what I can do. You'd better look for the killer.'
(WR-P-E-G-0000001942.p.414.s.1)

b.   *Maar dan zou   ik niet hier zijn om u    om hulp te  vragen.*
But    then would I   not here be   to you for help to  ask
*Zoek  liever die   moordenaar. Dat kan alleen als      we weten*
search rather that  killer          that can only   when    we know
*wat er    met Nico aan de hand is.*
what there with Nico on    the hand is
'But then I wouldn't be here to ask for your help. You'd better look for
that killer. That is only possible when we know what is going on with
Nico.'
                    (WR-P-E-G-0000002140.p.471.s.1)

89.  *Die jaagt  de konijnen  uit de pijpen, zodat     de vogels ze     kunnen*
that hunts  the rabbits    out the pipes  such_that the birds   can   grab
*pakken.*
them
'He hunts the rabbits out of the pipes, such that the birds can grab them.'
                    (WR-P-P-G-0000284233.p.4.s.8)


### 5.2.1.2      Borderline retained instances

Instances that appeared in newspaper or magazine headlines, such as (90), and
advertisements such as (91) were retained, because (i) these represent instances of
language usage that are in principle just as ecologically valid as any other form of
language usage; (ii) they appear in both variants, as in (90b) and (91b); (iii) we do
not want to exclude instances from the dataset simply because we suspect they may
be harder to model.


90.  a.   *Canadese tv zoekt    filmpjes     uit  oorlog.*
Canadian  TV searches little_films   from war
'Canadian TV looking for little films from war.'
                (WR-P-P-G-0000034643.head.1.s.1)

     b.   *VS zoekt    naar botten van  Saddam.*
US searches to    bones of    Saddam
'US looking for bones of Saddam.'
                (WR-P-P-G-0000014738.head.1.s.1)

91.   a.   *Ik  zoek  familie, kennissen      of  buren      van schilder en*
          I   search family  acquaintances or  neighbours of   painter
          *illustrator Roger De Ruyck, geboren in  Merelbeke op 12 juni  1918.*
          illustrator Roger De Ruyck  born      in  Merelbeke on 12 June 1918
          'I'm looking for family, acquaintances or neighbours of painter and
          illustrator Roger De Ruyck, born in Merelbeke on June the 12th 1918.'
                    (WR-P-P-G-0000212722.p.6.s.1)

      b.   *Ik zoek al      jaren naar een  vriendin     van wie    ik*
          I   search already years to   a     female_friend of    whom I
          *de achternaam niet  meer   weet.*
          the surname      no    longer know
          'I've been searching for years to a female friend whose name I've
          forgotten.'
                    (WR-P-P-G-0000440794.p.2.s.1)

### 5.2.2   *Verlangen* 'desire'

#### 5.2.2.1   Removed instances

Instances with an inanimate agent, as in (92a), hardly allow for the prepositional
variant. This might be due to the meaning difference between desire as 'demand'
and 'long for', hypothesized in Section 4.2.2.2.1. We propose that the meaning
difference is too dominant here, and hence remove these 33 instances from the
dataset (cf. Section 2.1).[55] Three of these instances did actually exhibit the
prepositional variant, but they had semi-animate inanimate agents, viz. body, soul
and eyes, as in (92b). These were also removed for consistency.

---

[55] The absence of a choice for the individual language user, e.g. the impossibility to choose
the prepositional variant, is a valid ground for exclusion following the definition of
alternations as an individual's choice points (cf. Section 2.1, e.g. Röthlisberger 2018b: 47–
65). Following the definition as researcher's setups, this exclusion could be justified by a
conservative research design. That is, in order for us to confirm the distinction between
'demand' and 'long for', we demand that SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-
CONSTRUCTION will still be positively correlated with the probability of the prepositional
variant, even when instances such as (92) are excluded (cf. Section 2.1).

92.   a.   *Van Lidth meende echter  dat  Eggens  de eigenschappen die*
           Van Lidth believed however that  Eggens  the properties      that
           *deze functie  verlangde miste.*
           this  function desired    missed
           'Van Lidth believed, however, that Eggens lacked the qualities that his
           function required.'
                    (WR-P-P-B-0000000064.p.137.s.7)

    b.   *Het hongerhormoon ghreline stijgt dan  weer  bij    slaaptekort*
           the   honger_hormone ghreline rises then again  with sleep_shortage
           *en   doet het lijf  naar meer  calorieën verlangen.*
           and  does the body to    more   calories   desire
           'The honger hormone Ghreline, then again, rises because of sleep
           deficiencies and makes the body desire more calories.'
                    (WR-P-P-G-0000205311.p.3.s.5)

One transitive instance where the verb *verlangen* 'desired' referred to an explicit question was removed, viz. (93).

93.   *Peter R. de Vries, die  ook  ter    plekke was en  het verhaal gehoord*
     Peter R. de Vries, that also at_the place  was and  the story     heared
     *had van Thomas van Luyn, wordt, evenals  de cabaretier,         aan*
     had of    Thomas van Luyn  was    just_like the cabaret_performer on
     *de telefoon   verlangd.*
     the telephone desired
     'Peter R. de Vries, who was also on site and had heard the story from
     Thomas van Luyn, is, just like the cabaret performer, is required at the
     telephone.'
                 (WR-P-P-G-0000053813.p.2.s.5)

Twelve transitive instances where the theme was referred to by the conjunctions *zoals* 'like' or *anders* 'otherwise', as in (94), were removed, as the prepositional variant did not seem possible.

94.  a.  *Bovendien wil   Washington een ander bewind in Bagdad,*
         Moreover  wants Washington a   other regime in Bagdad
         *en   niet alleen de terugkeer van de wapeninspecteurs, zoals*
         and not only the return  of the weapon_inspectors like
         *de VN  verlangen.*
         the US  desire
         'Moreover, Washington wants another regime in Bagdad, and not just
         the return of the weapon inspectors, like the UN desire.'
                    (WR-P-P-G-0000004003.p.3.s.3)

     b.  *Een ster om op aan te sturen,  anders   verlang ik niet.*
         a   star to on on to head_for otherwise desire  I   not
         'A star to head for, that is all I desire.'
                    (WR-P-P-G-0000000446.p.5.s.1)

Finally, we need to consider instances where the theme is a complement clause
without an enclosed antecedent, as in (95). In Dutch, such complement clauses may
be introduced by an anticipatory pronoun or pronominal adverb, as in (96a-b). The
anticipatory pronoun is said to be obligatory with adjectival complementizers, as
in (96c), while for the anticipatory pronominal adverb, this seems to depend upon
the verb. That is, anticipatory pronominal adverb is optional for *klagen* 'complain'
in (96b), but allegedly obligatory for *genieten* 'enjoy' in (96d).[56] In the cases where
the presence of the anticipatory pronoun or pronominal adverb is optional, this
alternation appears to be driven by lexical biases and semantic nuances (Broekhuis
and Corver: 645–647, 675–677), although we are unaware of an in-depth
alternation study on the matter.

   Now, it has been proposed that the prepositional variant of the verb *verlangen*
'desire' presents a case where the anticipatory pronominal adverb is optional (Van
de Velde 2014a: 340). If that is true, it is impossible to ascertain whether the
instances (95) constitute cases of the transitive or the prepositional variant. As
such, they had to be excluded from the dataset.

---

[56] Instances such as (96c-d) do certainly occur without the anticipatory pronoun or
pronominal adverb in the Sonar-corpus, however. Here are two examples: *Ik vind leuk dat
je op m'n liedje goed kan dansen en zingen.* 'I like that you can nicely dance and sing along
with my song' (WR-P-E-G-0000009588.p.100.s.2) and *…, maar ergens denk ik ook dat je niet
schuldig hoeft te voelen dat je even geniet dat je alleen bent.* 'but I also think that you
shouldn't have to feel guilty for enjoying that you are alone.' (WR-P-E-A-0005727223
.p.5.s.1). Perhaps the posited differences in grammaticality may be more accurately described
as lexical biases in the alternation between the presence versus absence of anticipatory
pronouns and pronominal adverbs.

95.   a.   *Ten   eerste verlangt de Unesco dat alle objecten die  zij op*
          at_the first   desires the Unesco that all   objects that she on
          *haar erfgoedlijst         zet ook beschermd worden door de*
          her world_heritage_list puts also protected   are       by   the
          *nationale  wetgeving.*
          national    legislation
          'First, the Unesco desires that all objects that she puts on her world
          heritage list, are also protected by national legislation.'
                     (WR-P-P-H-0000120768.p.3.s.1)

    b.   *Als    het seizoen stopt, verlang je    alweer... dat de eerste*
          when   the season  stops  desires  your again    that the first
          *koers  weer  in België  is.*
          race    again  in  Belgium is
          'When the season stops, you again desire... that the first race comes to
          Belgium once more.'
                     (WR-P-E-G-0000008418.p.100.s.1)

    c.   *Het gevolg       is dat iedereen en   niemand wel ergens*
          the  consequence  is  that everyone and  no_one part somewhere
          *blijk  geeft  van  een  nood om te ontsnappen, iedereen en*
          mark  gives  of   a    need to to escape    everyone and
          *niemand plotseling verlangt om binnen te  blijven.*
          no_one suddenly desire   to inside to stay
          'The consequence is that everyone and no one somewhere exhibits the
          need to escape, everyone and no one suddenly desires to stay inside.'
                     (WR-P-P-H-0000119867.p.3.s.4)

96.   a.   *Hij waardeert (het) dat je  elke   dag  op bezoek komt.*
          he    appreciates (it)  that  you  every  day  on  visit    comes
          'He appreciates it that you come for a visit every day.'

    b.   *Hij klaagt   (erover)    dat je   elke   dag  op bezoek  komt.*
          he complains (thereabout) that  you  every  day  on  visit    come
          'He complains that you come for a visit every day.'

    c.   *...dat Jan  *(het)  vervelend  vindt dat hij niet kan komen.*
          that Jan  *(it)    unpleasant finds  that  he  not  can  come
          '...that Jan regrets that he is not able to come.'
                 (taken over from Broekhuis and Corver: 768)

> d.  *Hij geniet \*(ervan) dat je elke dag op bezoek komt.*
>     he  enjoys \*(thereof) that you every day on visit  comes
>     'He enjoys it that you come for a visit every day.'

This exclusion may introduce a bias into our dataset, however. We currently do not know how the alternation between presence vs. absence of the anticipatory pronoun and pronominal adverb is determined. As such, it is possible that the anticipatory pronoun is more easily omitted than the anticipatory pronominal adverb. In fact, we may have found an indication that this is indeed the case in our dataset: we did not discover any instances where an anticipatory pronoun was explicitly expressed, such as (97a), while we did find many instances with an anticipatory pronominal adverb, like (97b).[57] In other words, had the transitive variant been used in (97b), even all other things being equal, the anticipatory pronoun *het* might have been omitted. As a result, (97b) would then have been removed from the dataset, and the dataset would be biased against the transitive variant. As such, we also removed instances like (97b) from the dataset, even though the choice of variant here is clear. In total, 54 instances where the theme is a complement clause without an enclosed antecedent, were removed from the dataset.

97.  a.  *Maar de jongere generatie  verlangt het om in België  zelf*
         but  the younger generation  desires  it  to in Belgium itself
         *begraven te  worden.*
         buried   to be
         'But the younger generation desires to be buried in Belgium itself.'

     b.  *Maar de jongere generatie  verlangt ernaar  om in België*
         but  the younger generation  desires  thereto to in Belgium
         *zelf begraven te  worden.*
         itself buried   to be
         'But the younger generation desires to be buried in Belgium itself.'
                 (WS-U-E-A-0000072402.p.1.s.3)

---

[57] However, this could also be interpreted as an indication that the anticipatory pronominal adverb is, in fact, obligatory for *verlangen* 'desire', and hence, that all instances such as (95) instantiate the transitive variant, where the anticipatory pronoun can remain unexpressed. In that case, there would be no need to exclude instances like (95) and (97) from the dataset.

Note that none of this pertains to instances where the theme is a complement clause with an enclosed antecedent, such as (98), as the variant is always clear there. Therefore, such instances were retained in the dataset.

98.  *Verlang nu   eens niet  naar wat je    mist, maar verlang nu   eens naar*
     desire   now once not   to    what you  miss but  desire   now  once  to
     *wat  je    hebt.*
     what you   have
     'For once, desire not what you miss, but desire what you have.'
                     (WR-P-P-B-0000000056.p.1027.s.7)

### 5.2.2.2     Borderline retained instances

Instances where *verlangen* seemed to mean 'demand' (99) or 'long for' (100) were retained in the dataset, because (i) this seemed to occur in both the transitive and prepositional variant, cf. (99)-(100), (ii) this actually concerns our central semantic hypothesis for this verb (see Subsection 4.2.2.2.1), (iii) we saw no reliable way of strictly delineating both meanings from one another for individual instances.

99.  a.  *Hij vindt dat   minister Remkes de  dreiging in  Nederland*
         he  finds that   minister Remkes the threat   in  the_Netherlands
         *ernstig     onderschat      heeft, en   verlangt op korte termijn*
         seriously   underestimated  has     and desires  on short  term
         *'harde   maatregelen'*
         firm     measures
         'He thinks that minister Remkes has seriously underestimated the
         threat in the Netherlands, and demands "firm measures" in the short
         term.'
                         (WR-P-P-G-0000131620.p.2.s.2)

     b.  *Belangrijk is  ook dat  men niet meteen      resultaten verlangt.*
         Important  is  also that one not  immediately results    desires
         'It is also important that one does not immediately desires results.'
                         (WR-P-P-G-0000693758.p.6.s.5)

    c.   *Maar de politiek verlangt nu naar scherpere maatregelen om*
        but   the politics desires now to   scharper measures    to
        *de inburgering te verbeteren.*
        the naturalization to improve
        'But politicians now desire scharper measures to improve
        naturalization.'
                (WS-U-E-A-0000000019.p.10.s.3)

    d.   *Ik verlang echt naar betere resultaten.*
        I   desire   really to   better results
        'I really desire better results.'
                (WR-P-P-G-0000714028.p.2.s.13)

100. a.   *Gewoon twee mensen die elkaar graag zien en een kind*
        just   two people   that each_other gladly see and a   child
        *verlangden.*
        desired
        'Just two people who loved each other and desired a child.'
                (WR-P-P-G-0000185386.p.2.s.4)

    b.   *En een knuffel verlangt toch iedereen.*
        and a   hug    desires still everyone
        'And everyone does desire a hug.'
                (WR-P-E-G-0000002046.p.273.s.1)

    c.   *Zij verlangde naar een kind.*
        she desired   to   a   child
        'She desired to a child'
                (WR-P-P-G-0000123915.p.24.s.6)

    d.   *Een groot deel van de kiezers verlangt wellicht naar de*
        a   large part of   the voters   desires possible to   the
        *geborgenheid van een leidersfiguur.*
        sense_of_security of   a   leader_figure
        'A large part of the voters probably desires the sense of security of a
        leader figure.'
                (WR-P-P-H-0000092985.p.25.s.1)

Instances where an additional prepositional constituent with *van* 'of' explicitly mentioned the person or object of whom something is being desired, as in (101), were retained in the dataset, even though these appeared predominantly, perhaps

even solely, in the transitive variant.[58] We have the following four reasons for this. First, the reasons why such instances prefer the transitive variant is likely because the person or object of whom something is desired is much more salient when the meaning of *verlangen* inclines towards 'demand' and away from 'long for'. Since we have decided to retain instances where the meaning of *verlangen* inclines towards 'demand', we want to be consistent with this choice and hence also retain instances like (101).

Second, these instances are rife with PP-attachment ambiguity, as in (102a-b). The prepositional variant is also used among instances with this ambiguities, as in (102c-d), and distinguishing between either attachments or even between ambiguous and unambiguous instances is ticklish at best, and practically impossible at worst.

Third, it seems inconsistent to exclude the instances where the person or object of whom something is desired is explicitly expressed, while retaining those where the participant could be easily added, which is the case in most transitive and prepositional instances. For instance, in (103), the constituents *van de winkelbedienden* 'of the shopping assistents', *van de Nationale Bank* 'of the Federal Reserve', *van hem* 'of him', *van zijn onderhandelingspartner* 'of his partner in negotiations' could be respectively added. Fourth, for the verb *peilen* 'gauge', such instances were also tolerated (see Subsection 5.2.3.2, examples in (119)).

101. a.   *Straks gaat hij weer dingen van mij verlangen.*
    Later goes he again things of me desire
    'Later, he'll start desiring things of me again.'
           (WR-P-E-G-0000008459.p.377.s.1)

   b.   *Het geduld dat Bush en Blair de wereldgemeenschap*
    the patience that Bush and Blair the global_community
    *destijds niet meer gunden, verlangen ze nu wel van*
    in_those_days not more granted desire they now PART of
    *diezelfde gemeenschap.*
    that_same community
    'The patience that Bush and Blair did not grant the global community, they themselves now do desire from that community.'
           (WR-P-P-G-0000014591.p.3.s.2)

---

[58] We write *perhaps even solely* because the prepositional instances with a *van*-participant in our dataset present cases of PP attachment ambiguities (see below). In one instance, this person of object was introduced by *aan* 'to' rather than *van* 'of'. This instance was also retained.

102. a. *Maar de individuele klant die voor ons staat, legt niet*
but   the individual   client that before us    stands lays not
*dezelfde burgerzin     aan de dag en   verlangt de medewerking*
the_same citizen_sense on  the day and   desires the cooperation
*van de bank om zijn erfenis    of vermogen zoveel     mogelijk*
of  the bank to his   inheritance or capital   so_much possible
*voor de fiscus   af   te schermen.*
for  the taxman PART to  screen
'But the individual client that stands before us, does not exhibit the
same sense of public responsibility en desires the cooperation of/from
the bank to screen off his inheritance or his capital from the taxman.'
(WR-P-P-H-0000043048.p.6.s.3)

b. *Schulz liet gisteravond     weten geen persoonlijke excuses  van*
Schulz let yesterevening   know no   personal      apologies of
*Berlusconi meer  te verlangen.*
Berlusconi more  to desire
'Schulz announced yesterevening not to desire anymore personal
apologies of/from Berlusconi.'
(WR-P-P-G-0000064428.p.5.s.2)

c. *Ik verlangde naar nieuwe berichten van Alcaeus.*
I desired   to  new    messages of  Alcaeus
'I desired new messages of/from Alcaeus.'
(WR-P-P-B-0000000170.p.730.s.5)

d. *Bij Renault wordt er binnenskamers     naar betere banden*
with Renault is  there behind_closed_doors to   better tires
*van Michelin verlangd.*
of   Michelin desired
'Many within Renault desire, behind closed doors, better desires
of/from Michelin.'
(WR-P-P-G-0000194939.p.5.s.1)

103. a. *Zo'n  man verlangt naar kleine dingen: dat  hij hartelijk wordt*
such_a man desires to    small thing    that he cordially is
*ontvangen;...*
received
'Such a man desires little things: that he is cordially received;...'
(WR-P-P-H-0000015490.p.7.s.4)

b.  *Ook al      verlangden misschien velen naar een  verlaging met*
    also already desired    perhaps   many to  a    reduction with
    *75 basispunten.*
    75 base_points
    'Even though many perhaps desire a reduction of 75 base points.'
            (WR-P-P-H-0000100042.p.12.s.3)

c.  *Vrouwen die  seksuele roofdieren zijn, die  net  als  de hongerige*
    women   that sexual   predators are  that just like the hungry
    *leeuw slechts één  ding   verlangen: hem verslinden en   ontdoen*
    lion  only    one  thing  desire     him devour    and  strip
    *van zijn viriliteit.*
    of  his  virility
    'Women that are sexual predators, that, just like the hungy lion, only
    desire one thing: to rip him apart and strip him of his virility.'
            (WR-P-P-H-0000175075.p.6.s.10)

d.  *De minister verlangde aanvankelijk een  verlaging van 5 procent.*
    the minister desired    initially     a    reduction of  5 percent.
    'The minister initially desired a reduction of 5 percent.'
            (WR-P-P-G-0000088765.p.2.s.3)


Instances where the theme was animate were also retained, because these appeared
both in the transitive and prepositional variant, as in (100a,c) and (104).


104. a.  *Het is een  stad voor hardwerkende, succesvolle mensen, met*
         it  is a    town for  hard_working  successful  people  with
         *harde vrouwen die  sterke mannen verlangen, niet kindvriendelijk.*
         hard  women   that strong men    desire      not  child_friendly
                 (WR-P-P-G-0000584267.p.11.s.2)

b.  *Een club als  Aartrijke kan  nooit iemand   verlangen die*
    a   club like Aartrijke can  never someone  desire    that
    *financieel niet haalbaar is.*
    financially not  feasible is
    'A club like Aartrijke can never desire someone who is financially not
    feasible.'
            (WR-P-P-G-0000584267.p.11.s.2)

c. *PSV verlangt naar Van Persie.*
PSV desires to Van Persie
'PSV desires Van Persie.'
(WR-P-P-G-0000163888.head.1.s.1)

d. *Ik verlangde naar hem.*
I desired to him
'I desired him.'
(WR-P-P-B-0000000021.p.1152.s.2)

Prepositional instances with the causative auxiliary *doen* 'do', as in (105) were retained, even though this auxiliary did not occur in the transitive variant. The reason for retention is that based on the author's intuitions, the transitive variant was possible here.

105. a. *Wat zou mij dan naar de nieuwe versie kunnen doen*
whatwould me then to the new version can do
*verlangen?*
desire
'What would then make me desire the new version?'
(WR-P-P-H-0000132311.p.2.s.4)

b. *Zo gaf hij zijn personeel een prima gelegenheid om meer met*
so gave he his employeesa prime opportunity to morewith
*hun auto te kunnen pronken, wat ook hun buren en*
their car to can flaunt wat also their neighbours and
*kennissen naar een wagen zou doen verlangen.*
acquaintance to a car would do desire.
'In that way, he gave his employees a prime opportunity to flaunt their car some more, which would also make their neighbors and acquaintances desire a car.'
(WR-P-P-H-0000017029.p.2.s.4)

Finally, transitive instances where the theme starts with *niets liever dan* 'nothing other than' as in (106a-b) were also retained, because the prepositional variant is still possible, at least certainly under the condition that another comparative adverb with a similar meaning is used, as in (106c-d).

106. a. *En    het koninkrijk verlangde niets     liever dan een eind aan*
      and the kingdom    desired    nothing rather than an   end  to
      *het    geweld.*
      the   violence
      'And the kingdom desired nothing else than an end to the violence.'
           (WR-P-P-B-0000000065.p.1294.s.5)

   b. *'De radeloze, redeloze, reddeloze      burger van  vandaag vindt*
      the  desperate irrational irrecoverable citizen of    today   finds
      *het welletjes      en    verlangt niets     liever dan dictatuur.*
      it  quite_enough and  desires  nothing rather than dictatorship
      'The desperate, irrational citizen, lost beyond hope, has had quite
      enough and desires nothing else than dictatorship.'
           (WR-P-P-H-0000152981.p.2.s.12)

   c. *Naar niets     verlangt zij   heviger       dan naar vruchtbaarheid*
      to   nothing desires  she   more_intensely than to    fertility
      *en    bloei:...*
      and  blossom
      'She desires nothing more intensely than fertility and blossom:...'
           (WR-P-P-G-0000066187.p.7.s.3)

   d. *Zij die altijd zo hoog opgaf   over  taal,      kon   nu   naar*
      she who always so higly praised about language could now  to
      *niets    anders verlangen dan  naar Dario's  handen op haar lijf*
      nothing else   desire     than to    Dario's hands  on her  body
      *wanneer ze  thuiskwam.*
      when     she home_came
      'She, who always spoke so highly of language, could now desire
      nothing else than Dario's hands on her body when she came home.'
           (WR-P-P-B-0000000145.p.904.s.2)

## 5.2.3    *Peilen* 'gauge'

### 5.2.3.1    Removed instances

We first list the cases that clearly had to be removed. These were the following.

- 36 transitive instances where the theme was the participant being asked
  something, as in (107).

- 8 transitive instances where the theme was the result of the action of gauging, as in (108).
- 5 transitive instances where the theme was a fluid, as in (109).
- 2 transitive instances where the theme was being ranked on a certain level, as in (110).
- 1 transitive instance where the meaning of *peilen* was 'to detect', viz. (111).
- 2 instances where the theme was a complement clause without an enclosed antecedent, such as (112), as was done for *verlangen* 'desire'.

107. *Persbureau    Reuters had daags ervoor    22 vooraanstaande economen*
new_agency Reuters had day    therefore 22 prominent    economics
*gepeild en   iedereen voorspelde een verlaging van ten minste 100*
gauged and everyone predicted a    reduction of at least 100
*of zelfs 200 basispunten.*
of even 200 basic_points
'The news agency Reuters had probed 22 prominent economics and
everyone predicted a reduction of at least 100 or even 200 basic points.'
                    (WR-P-P-G-0000078773.p.13.s.3)

108. *Maar een nieuwe studie peilt    veel    minder enthousiasme voor de*
but a    new    study gauges much less    enthusiasm for the
*doodstraf    dan gedacht.*
dead_penalty than though
'But a new study gauges a lot less enthusiasm about the dead penalty than
anticipated.'
                    (WR-P-P-G-0000058286.p.1.s.2)

109. *Kon   ik nog rustig doorrijden en   elk uur olie peilen?*
could I    still calmly drive_on and each hour oil    gauge
'Could I still calmly drive on and gauge the oil every hour?'
                    (WR-P-P-B-0000000210.p.623.s.3)

110. *In vergelijking met de vorige    peiling verliest CD&V echter   1,3*
in comparison with the previous poll    loses    CD&V however 1.3
*procent, terwijl de partij nu    samen   met de N-VA wordt gepeild.*
percent while the party now together with the N-VA is    gauged
'In comparison to the previous poll, CD&V now loses 1.3% percent, while
the party is now estimated together with the N-VA'
                    (WR-P-P-G-0000222629.p.4.s.2)

111. *Elke   keer  dat  het AT daarna    wéér  een uitzending   peilt,  is een*
     every  time  that  the AT thereafter again  a    transmission gauges is a
     brief   voldoende om 2250 euro boete  op   te leggen.
     letter  sufficient to 2250 euro fine   PART to impose
     'Every time that the AT detect a transmission thereafter, a letter suffices to
     impose a fine of 2250 euro.'
                    (WR-P-P-G-0000102715.p.5.s.3)

112. *Volgend onderwerp... Voor de P peilt  Bruno of     Phara  politici*
     Following topic          for  the P gauges Bruno whether Phara  politicians
     *de poepers geeft.*
     the shits     gives
     'Next topic... for the P, Bruno gauges whether Phare gives the shits to
     politicians.'
                    (WR-P-E-G-0000006691.p.213.s.1)


Furthermore, 21 transitive instances with an animate theme that was being
evaluated by the agent, as in (113a), were removed from the dataset. The
prepositional variant does appear with animate themes, but only in collectives, as
in (113b), or as 'the (wo)man (in/behind)', as in (113c).[59] The meaning difference
between instances such as (113a) and those like (113b-c) was judged to be too
dominant to be retained in the dataset.


113. a.   *De burgemeester peilde    hem met  kille ogen.*
          the major            gauged  him with cold eyes
          'The major gauged him with cold eyes.'
                    (WR-P-P-B-0000000136.p.517.s.1)

     b.   *Om een correcter      beeld   te schetsen van de werkloosheid*
          to   a   more_correct picture to sketch  of   the unemployment
          *(...) becijferde het Vlaams Economisch Verbond (VEV) de*
               calculated the Flemish Economic    union    (VEV) the
          *werkloosheidsgraad op basis van  een NIS-enquête die  naar de*
          unemployment_rate on basis of   a    NIS-poll      that to  the
          *werkwillige        beroepsbevolking peilde.*
          willing_to_work work_force         gauged
          'To sketch a more correct picture of the unemployment, the Flemish
          Economic Union (FEU) calculated the rate of unemployment on the
          basis of a NIS-poll that gauged the work force that is willing to work.'
                    (WR-P-P-H-0000024416.p.3.s.1)

---

[59] These instances were retained in the dataset.

    c.    *Le Soir-sportjournalist  Stéphane Thirion peilde    naar de  mens*
          Le Soir-sport_journalist  Stéphane Thirion gauged   to    the human
          *Merckx  in  'Eddy Op en Top'.*
          Merckx  in  'Eddy Op en Top.'
          'Sports journalist Stéphane Thirion gauged the man Merckx in 'Eddy
          Op en Top'.
                   (WR-P-P-G-0000370769.p.1.s.2)

Two reflexive instances were removed from the dataset, viz. (114a-b), because these did not appear in the prepositional variant.


114.  a.    *De diepte van  hun ontroering  laat zich nauwelijks  peilen,...*
          the depth of   their emotion    lets itself hardly     gauge
          'The depth of their emotion is nearly impossible to judge.'
                   (WR-P-P-H-0000120510.p.4.s.7)


    b.    *Daarom   laten diepe  zielenroerselen bij   schermers zich*
          therefore let   deep   inner_emotions with fencers     themselves
          *moeilijk peilen.*
          hard     gauge
          'That why the deep inner emotions of fencers are hard to gauge.'
                 (WR-P-P-G-0000124480.p.1.s.2)


Ten prepositional instances where the agent seeks to determine the highest in some ranking, such as (115), were removed from the dataset. The reason is that the transitive variant would only be possible in such instances given a clearly distinct meaning. For instance, *Het toeristische magazine Outside peilde de mooiste bergwegen* would mean 'the tourist magazine Outside evaluated the most beautiful maintain roads', where the set of most beautiful maintain roads would have been determined beforehand. In fact, such instances did not appear with the transitive variant.


115.  a.    *Het toeristische magazine Outside peilde,  in  samenwerking*
          the  tourist      magazine Outside gauged  in  cooperation
          *met verscheidene automobilistenclubs, naar de  mooiste*
          with several        drivers_clubs       to   the most_beautiful
          *bergwegen     ter    wereld.*
          mountain_roads  in_the world
          'The tourist magazine Outside sought to determine, in coorperation
                 (WR-P-P-H-0000103163.p.1.s.1)

b.  *In 2002 peilde de Nederlandse ouderenorganisatie ANBO bij*
in 2002 gauged the Netherlandic senior_society ANBO with
*180.000 van haar leden naar de ideale telefoon voor*
180,000 of her members to the ideal phone for
*vijftigplussers.*
people_over_fifty
'In 2002, the Netherlandic senior-oganization ANBO sought to
determine, among 180,000 of her members, the ideal phone for people
over fifty.'
(WS-U-T-B-0000000926.p.3.s.1)

### 5.2.3.2    Borderline retained instances

Instances with an inanimate theme that was being judged, were retained in the
dataset, because these appeared both in the transitive and prepositional variant, as
in (116).

116. a.  *Door dat raampje probeerde ik het weer te peilen*
through that little_window tried I the weather to gauge.
'Through that little window, I tried to gauge the wheater.'
(WR-P-P-B-0000000190.p.212.s.4)

b.  *Wie kan de uitgestrektheid van de wolken peilen, (...) ?*
who can the vastness of the clouds gauge
'Who can gauge the vastness of the clouds?'
(WR-P-P-B-0000000427.p.2785.s.1)

c.  *Rothko's hagiografen uit de jaren vijftig strooiden (...) om*
Rothko's hagiographers out the years fifty sprinkle to
*naar de mythische en spirituele dimensie van de kleurenvelden*
to the mythical and spiritual dimension of the color_fiels
*te peilen.*
to gauge
'Rothko's hagiographers sprinkled (...) to gauge the mythical and
spiritual dimension of the color fields.'
(WR-P-P-H-0000058181.p.25.s.7)

d.   *We kunnen door de vergelijking van mensen en   apen wel naar*
   we   can   by   the comparison of   humans and   apes PART to
*de aard   van het taalvermogen          of indicatoren voor*
   the nature of   the linguistic_competence or   indicators   for
*agressie    peilen.*
   aggression   gauge
‘By comparing humans and apes, we can gauge the nature of linguistic
competence of indicators of aggression.’
       (WR-P-P-G-0000348817.p.11.s.5)

Instances with an inanimate theme that was asked for, were retained in the dataset, because these appeared both in the transitive and the prepositional variant, as in (117).

117. a.   *Hij keerde zich   tot zijn raadslieden en   peilde   kort    hun*
      he   turned himself   to his   counselors and   gauged shortly   their
*antwoord, dat unaniem    was.*
      answer    that unanimous was
‘He turned to his counselors, and shortly gauged their answer, which was unanimous’
      (WR-P-P-B-0000000054.p.719.s.2)

b.   *Net voor   ze   de eerste kasseistrook op doken, peilde   het*
   just before   they the first    cobble_strip on dove   gauged the
*Vlaamse supertalent de mening van zijn Nederlandse collega*
   Flemish supertalent the opinion of   his   Netherlandic college
*over   de Pro Tour.*
   about   the Pro Tour
‘Just before they reached the first strip of cobble stones, the Flemish supertalent gauged the opinion of his Netherlandic college about the Pro Tour.’
      (WR-P-P-G-0000042916.p.2.s.2)

c.   *De Standaard belde   daarom    naar VRT-woordvoerster Diane*
   De Standaard called   therefore to    VRT-spokesperson Diane
*Waumans en   peilde   naar de nieuwe plannen.*
   Waumans and   gauged   to   the new    plans
‘Therefore, De Standaard called VRT-spokesperson Diane Waumans and gauged the new plans.’
      (WR-P-E-C-0000008978.p.4.s.1)

d.    *'What's in  a name' dachten we en   peilden  naar jullie mening.*
      What's in  a name  thought  we and  gauged to    your opinion
      '"What's in a name" we thought, and we gauged your opinion.'
            (WR-P-E-C-0000009624.p.1.s.4)

Instances with a theme which was searched for, but which was  not the highest in
some ranking, were retained in the dataset, because these appeared both in the
transitive and the prepositional variant (118).

118. a.    *Ze    vraagt wetenschappelijk onderzoek om de  oorzaken  van  de*
            she   asks   scientific          research to the causes    of   the
            *scheeftrekking  te  peilen.*
            distortion         to  gauge
            'She asks scientific research to gauge the causes of the distortion.'
                  (WR-P-P-G-0000235681.p.4.s.6)

   b.    *Het Amerikaanse onderzoeksbureau  Pew Research Center  peilt*
         The American     research_center    Pew Research Center  gauges
         *al     enige jaren de vermeende botsing der    beschavingen.*
         already some  years the alleged    clash   of_the civilizations
         The American research center Pew Research Center already gauges
         the alleged clash of civilizations for some years.'
               (WR-P-P-H-0000153593.p.2.s.1)

   c.    *Andersen Consulting peilde   in een  gedetailleerde studie  naar de*
         Andersen Consulting gauged  in a     detailed        study  to   the
         *oorzaken  van  hun  succes.*
         causes     of    their success
         'Anderse Consulting gauged the causes of their success in a detailed
         study.'
               (WR-P-P-H-0000031204.p.19.s.2)

   d.    *De tentoonstelling over  de sfinx  in het ING in  Brussel  peilt*
         the exhibition         about the Sphinx in the ING in  Brussels gauges
         *naar de oorsprong van  het hybride wezen   en     volgt    de*
         to    the origin     of   the hybrid   creature and  follows  the
         *stilistische evolutie.*
         stylistic    evolution.
         'The exhibition about the Sphinx in the ING in Brussels gauges the
         origin of the hybrid creature and follow the stylistic evolution.'
               (WR-P-P-G-0000260179.p.8.s.2)

Instances where an additional prepositional constituent explicitly mentioned the person of whom something was being gauged, were retained in the dataset, because these appeared both in the transitive and the prepositional variant (119).

119. a. *JP Morgan ging gisterochtend bij financieel directeur Ray*
      JP Morgan went yestermorning with financial director Ray
      *Stewart van Belgacom de interesse peilen (...)*
      Steward of Belgacom the interest gauge
      'JP Morgan went to gauge the interest of financial director Ray Stewart of Belgacom...'
            (WR-P-P-G-0000229179.p.1.s.5)

    b. *Het Eén-magazine Koppen zendt het gesprek uit en*
      The Eén-magazine Koppen broadcasts the comversation part and
      *ging de reacties peilen van Vlaamse betrokkenen.*
      went the reactions gauge of Flemish those_involved.
      'The Eén-magazine Koppen broadcasts the conversation and went to gauge the reactions of Flemings that were involved.'
            (WR-P-P-G-0000262161.p.2.s.3)

    c. *De Mandel peilt naar interesse bij handelaars voor sociaal*
      De Mandel gauges to interest among traders for social
      *woonproject.*
      housing_project
      'De Mandel gauges interest among traders for social housing project.'
            (WR-P-P-G-0000397611.head.1.s.1)

    d. *Rudi Vranckx peilde in de stad Lahore naar de reacties en*
      Rudi Vranckx gauged in the city Lahore to the reactions and
      *verwachtingen van vrouwen.*
      expectations of women
      'Rudi Vranckx gauged the reactions and expectations of women in the city of Lahore.'
            (WS-U-E-A-0000096261.p.1.s.5)

## 5.2.4    *Zoeken* 'search'

### 5.2.4.1    Removed instances

The following instances of idiomatic expressions were removed from the dataset

- 386 transitive instances of *er iets achter zoeken* with the figurative meaning 'read something into it', as in (120a). Sentences that include an adjunct with *achter* in the literal meaning were retained, as in (120b).
- 117 transitive instances of *(zelf) gezocht* and *het (gaan) zoeken* with the figurative meaning 'it's their own fault' or 'they're responsible themselves' or 'they took unwarranted risks', as in (121a-b), were also removed. Again, instances with a literal meaning were retained, as in (121c-d).
- 874 transitive instances of *heeft te zoeken*, lit. 'has to search', *gaan zoeken*, lit. 'go search' and *komen zoeken*, lit. 'come search' with the figurative meaning 'have (no) bussiness' such as (122).
- 837 transitive instances of *het zoeken in*, lit. 'search it in', with the figurative meaning 'focus on', as in (123).
- 18 transitive instances of *niet weten waar het ze het moeten zoeken*, lit. 'don't know where they should search for it', with the figurative meaning 'they panicked', as in (124).
- 34 transitive instances of the expressions *soort zoekt soort* or *ons zoekt ons* 'birds of a feather flock together', *verhaal zoeken bij* 'settle things with', lit. 'search story with', and *de mosterd zoeken in* 'search inspiration in', lit. 'search for the mustard in', as in (125).


120. a.   *Moeten we daar iets        achter zoeken?*
          Must    we there something   behind search
          'Should we read something into that?'
                   (WS-U-E-A-0000188908.p.1.s.16)

     b.   *Anna Carla zocht    de rugleuning van de stoel achter zich,   en*
          Anna Carla searched the back        of    the chair behind herself and
          *ging zitten.*
          went sit
          'Anna Carla searched the back of the chair behind her and sat down.'
                   (WR-P-P-B-0000000038.p.1749.s.3)

121. a.    *Die rel        heeft hij duidelijk  gezocht. Ratzinger citeerde niet*
            that commotion has  he  clearly     sought  Ratzinger cited      not
            *zomaar   een   middeleeuwse  keizer.*
            just      a     medieval       emperor
            'He clearly intended to raise commotion. Ratzinger did not just cite
            any random medieval emperor.'
                        (WR-P-P-G-0000258037.p.17.s.1)

     b.    *We stonden 0-1 voor  en    we speelden  met elf     tegen     tien.*
            we  stood   0-1 before and  we played     with eleven against  ten
            *Dan moet  je    het niet gaan zoeken.*
            then should you   it  not  go   search
            'We were ahead 1-0 and we were playing eleven versus ten. Then you
            should not take any unwarranted risks.'
                        (WR-P-P-G-0000315406.p.5.s.3)

     c.    *Ik heb heel eerlijk     tegen    mezelf gezegd... Van Goethem, jongen,*
            I   have very honestly against  myself told      Van Goethem  boy
            *wat  er    ook gebeurt... Waar je    nu    behoefte aan hebt, dat is*
            what there also happens   where you  now   need     on   have  that is
            *aan gezelschap. Maar je    moet het wel zelf      gaan zoeken.*
            on   company     but  you  must it  part yourself  go   search
            'I've told myself very honestly... Van Goethem, son, whatever
            happens... what you need now, is company. But you'll have to seek it
            out yourself.'
                        (WR-P-E-G-0000003379.p.211.s.1)

     d.    *Veel   water kunnen ze   niet meenemen. Daarom    moeten  ze*
            much  water can    they not  take_with  therefore must    they
            *het gaan zoeken, zoals in  een  opgedroogde  rivierbedding.*
            it  go   research like  in  a    dried_out     riverbed
            'They cannot bring a lot of water whit them. That's why they have to
            go look for it, such as in a dried out riverbed.'
                        (WR-P-E-G-0000007856.p.37.s.1)

122. a.    *En  dat  terwijl ze    daar niks      hebben te zoeken.*
            and that while  they  there nothing  have    to search
            'And that, while they don't have any business being there.'
                        (WR-P-P-G-0000014386.p.5.s.9)

b.  *Claustrofoben    kunnen in New York   weinig gaan zoeken.*
Claustrophobics  can      in New York  little   go    search
'Claustrophobics don't have much to do in New York.'
(WR-P-E-G-0000005329.p.106.s.1)

c.  *Ja... Anders    was ik bij  jou  niks     komen zoeken.*
Yes otherwise  was I   with you  nothing  come   search
'Yes... Otherwise I would have no reason to come to you.'
(WR-P-E-G-0000002702.p.180.s.1)

123.  *Wij zoeken  het vooral     in banen in het bedrijfsleven,    waarbij*
we  search  it primarily  in jobs   in the private_sector whereby
*jongeren leren  èn    werken.*
youngster learn   and   work
'We mainly focus on jobs in the private sector, whereby youngsters learn ánd work.'
(WR-P-P-G-0000063538.p.8.s.4)

124.  *De  Engelsen wisten niet waar  ze    het zoeken   moesten.*
The  English    knew  not  where they  it  search    must
'The English didn't know what to do'
(WR-P-P-H-0000154540.p.5.s.5)

125.  a.  *Soort zoekt duidelijk soort, lijken de foto's te zeggen.*
sort    search clearly sort,   seems  the picture to say
'Birds of a feather flock together, seems to be expressed by the pictures.
(WR-P-P-G-0000031563.p.4.s.6)

b.  *U    kunt verhaal zoeken  bij   de rechter.*
you  can  story    search   with the judge
'You can settle things with the judge.'
(WR-P-P-H-0000091549.p.12.s.4)

c.  *Beide kunstenaars waren 'romanisten',  die   de mosterd in Italië*
both   artists      were  romanists    that the mustard in Italy
*zochten.*
searched
'Both artists were 'romanists', that searched for inspiration in Italy.'
(WR-P-P-H-0000122749.p.6.s.3)

### 5.2.4.2    Borderline retained instances

Instances of the expressions *(naar) een dubbele bodem zoeken* 'look for a hidden meaning', lit. 'search for a double bottom', *(naar) een speld/naald in de hooiberg zoeken* 'searching for a needle in a haystack', *(naar) een stok zoeken om mee te slaan* 'look for an excuse to attack someone', lit. 'look for a stick to beat' were retained, because these regularly appeared in both variants. Instances of *(naar) een dak boven ons hoofd zoeken* 'look for a place to stay', lit. 'look for a roof above our heads', *(naar) een gulden middenweg zoeken* 'search for a happy medium', lit. 'search for a golden middle road', *(naar) een luisterend oor zoeken* 'search for someone to talk to', lit. 'search for a listening ear' and *(naar) spijkers op laag water zoeken* 'split hairs', lit. 'search for nails on low water' only appeared in one variant, but were nonetheless retained. This was done because this was suspected to be simply due to their low frequencies.

Instances where the theme argument was a criminal or suspect, as in (126), were retained because these appeared in both variants.

126.  a.   *Zes Marokkanen worden ervan    verdacht de daders      te*
            six  Moroccans   are      thereof suspected the perpetrators to
            *zijn van de    aanslagen in Madrid. Vijf van hen worden nog*
            be  of  the  attacks   in Madrid  five of   them are     still
            *gezocht.*
            sought
            'Six Moroccans are suspected of being the perpetrators of the terrorist attacks in Madrid. Five of them are still being sought.'
                     (WR-P-P-G-0000162189.p.5.s.2)

      b.   *Want het is op gezag      van Justitie dat naar de vijf wordt*
            for   it  is on authority of   Justice that to   the five is
            *gezocht.*
            sought
            'For it is on the authority of the court that the five are being sought.'
                     (WR-P-P-H-0000152511.p.7.s.7)

Instances of modal *zijn te* 'be to' plus infinitival *zoeken* 'to search' were retained, as in (127a). The reasons are that (i) they do appear in the prepositional variant, as in (127b), if only scarcely; (ii) the same meaning is regularly expressed in the prepositional variant with only slight changes to the sentence structure, as in (127c-d). In the same vein, instances of *zijn* 'be' plus *het zoeken*, like (127e), were also retained.

127.  a.  *De keerzijde hoort  er   wel bij: enige complexiteit is  ver te*
          the downside belongs there part with any    complexity   is  far to
          *zoeken.*
          search
          'There is also a downside to this: any complexity is hard to find.'
                    (WR-P-P-G-0000209846.p.4.s.1)

      b.  *...zodat   ze  online zijn te lezen en   binnen de teksten  naar*
          such_that they online are  to read  and  within the texts       to
          *trefwoorden is te  zoeken.*
          headwords  is to  search
          '...such that they can be read online and that it is possible to search
          headwords within the texts.'
                    (WR-P-P-G-0000097160.p.2.s.3)

      c.  *Naar subtiliteit kun je   lang zoeken  in de Latijnse literatuur.*
          to   subtlety   can you long search in the Latin    literature
          'Subtleties are hard to find in Latin literature.'
                    (WR-P-P-H-0000149983.p.1.s.4)

      d.  *Naar de    reden  waarom de goudmijnaandelen zulke zware klappen*
          to   the reason why   the gold_mine_stocks  such heavy beatings
          *hebben  gekregen, hoeft men niet lang te  zoeken.*
          have   got    has one not  long to search
          'The reason why the stocks of gold mines have received such heavy
          beatings, is not hard to find.'
                    (WR-P-P-H-0000032366.p.3.s.1)

      e.  *Naar koffie is het  ver zoeken.*
          to   coffee is it    far search
          'Coffee is hard to find.'
                    (WR-P-P-G-0000388032.p.1.s.4)


Instances with an extra-thematic reflexive argument were also retained, as in
(128a-b). The reason is that the same meaning can easily be, and is regularly,
expressed in the prepositional variant by simply dropping the reflexive argument
(128c). Like (128b), instances of the theme *weg* 'way' with a possessive pronoun,
as in (128e) are also retained, as they alternate in the same vein with instances like
(128f).

128. a. *Wie wapens    heeft maar geen vijand, zoekt    zich      een  nieuwe*
      who weapons  has   but  no   enemy searches oneself  a    new
      *vijand.*
      enemy
      'Who has weapons, but no enemy, finds oneself a new enemy.'
              (WR-P-P-H-0000083142.p.3.s.4)

   b. *Ze   zoeken  zich         een  weg naar het Zuiden, waar  ze*
      they  search  themselves a  way  to   the South   where they
      *hopen dat   het beter  is.*
      hope   that  it  is     better
      'They are looking for a way to the South, where they hope things are
      better.'
              (WS-U-E-A-0000188223.p.1.s.2)

   c. *Die groep zoekt    altijd    naar nieuwe  vijanden*
      that group searches always  to   new      enemies
      'That group is always looking for new enemies.'
              (WR-P-P-G-0000250738.p.5.s.3)

   d. *We  zoeken  naar mogelijke vijanden voor Bond.*
      we  search  to    possible  enemies for  Bond
      'We search for possible enemies for Bond.'
              (WR-P-P-G-0000011173.p.6.s.2)

   e. *Elke  computer zocht    zijn weg via  een andere computer om*
      every computer searched its  way  via  an  other  computer to
      *zijn informatie  door te geven.*
      his  information PART to  give
      'Every computers searched its way via another computer to pass on its
      information'
              (WR-P-P-H-0000123548.p.22.s.3)

   f. *… maar Aquiuss zocht    met zijn ogen naar een  andere weg om*
         but  Aquiuss searched with his  eyes to   an   other  way  to
      *op het hoogste dak  te  komen.*
      on the highest  roof to  come
      'But Aquiuss searched with his eyes for another way to get to the
      highest roof.'
              (WR-P-P-B-0000000026.p.2316.s.1)

Instances where the top in some ranking is being sought, and typically awarded the title expressed in the theme participant, are also retained in the dataset, as they appear in both variants (129).

129. a.  *Ze    zochten de meest sexy Rode   Duivel.*
         they  searched the most   sexy Red   Devil
         'They looked for the most sexy Red Devil.'
                    (WR-P-E-G-0000007181.p.119.s.1)

     b.  *Verder      zoekende organisatoren naar de Beste  Speelster  en*
         furthermore search the organisers     to   the Best   Player      and
         Vrouwencoach, Scheidsrechter en    Rookies van  het Jaar.
         Women_coach  Referee          and  Rookies of   the Year
         'Furthermore, the organizers looked for the Best Player and Women's coach, Referee and Rookies of the Year.'
                    (WR-P-P-G-0000360508.p.1.s.8)

## 5.2.5    Motoric verbs

### 5.2.5.1    Removed instances

The motoric verbs are the verbs that express a movement of a body part, viz. *grijpen* 'grab', *graaien* 'grasp', *grabbelen* 'scramble', *happen* 'snap' and *schoppen* 'kick'. The following instances of the verb *grijpen* 'grab' were removed from the dataset. The corresponding instances would also have been removed for the near-synonymous verbs *graaien* 'grasp' and *grabbelen* 'scramble', but these did not always occur, either because those verbs are a lot less frequent or because they are simply not used in such ways. When they did occur and were removed, this is explicitly mentioned below. In the same way, when corresponding instances of the verbs *happen* 'snap' and *schoppen* 'kick' were removed, this is also explicitly indicated.

- 800 transitive instances that describe a collision with a vehicle, as in (130), as these did not appear in the prepositional variant.
- 23 transitive instances that have a force of nature as their agent, such as (131).
- 336 transitive instances where *grijpen* means 'to catch up with', as in (132a), 5 where it means 'to hug', as in (132b), 92 where it means 'to take into costudy', as (132c), 10 where it means 'to take with you', as in

(132d), 16 where it means 'to understand' (132e) or 52 where it means 'to conquer', 'to take over' or 'to win', like (132f).

- 1 transitive instance of *moed grijpen* 'take hearth', viz. (133).
- 2 transitive instances of *het woord grijpen*, meaning 'take the floor to dominate a discussion', as in (134).
- 3 transitive instances where the theme marks a quantity, as in (135).
- 262 transitive instances that involve a mental or emotional captivation, such as (136a), as well as 4 prepositional instances that also express such a captivation and have *hart* 'hearth' as theme lemma, such as (136b).
- 2258 instances that involve the grabbing of an opportunity, as in (137), even though one such instance did exhibit the prepositional variant, viz. (137c). This was also removed.
- 336 instances that describe an animate being attacking by another animate being, as in (138). These instances predominantly exhibited the transitive variant, although the two instances in (138b-c) did exhibit the prepositional variant. Still, (138b) is a quote from Franz Kafka, and (138c) is religious language, so it was decided that these were exceptional instances and did not suffice for such instances to be retained. As such, instances like (138) were removed. 1 such instance were also removed for the near-synonymous verbs *graaien* 'grasp', as in (138d).
- 1764 instances where the theme participant is being grabbed by a body part or a piece of clothing, such as (139). This includes seemingly prepositional instances where the participant being grabbed is expressed in a possessive determiner to the body part or piece of clothing, as in (139b). 2 such instances were also removed for the near-synonymous verbs *graaien* 'grasp' (139c), as well as 10 such instances for *grabbelen* 'scramble' (139d), 9 such instances for *happen* 'snap' (139e) and 162 such instances for *schoppen* 'kick' (139f).
- 18 prepositional instances where it is clear from the context that the act of grabbing failed, as in (140a). 7 such instances were also removed for *graaien* 'grasp' (140b), 2 such instances *grabbelen* 'scramble' (140c) and 3 such instances for *happen* 'snap' (140d).
- 1248 prepositional instances where *grijpen* can be translated as 'use' or 'choose for', such as (141a-b). 5 transitive instances did appear to have those meaning, as in (141c-d), yet these were assumed to be exceptional in some way. They were also removed for consistency.
- 68 prepositional instances such as (142) that mean 'begin or decide to drink', as these did not appear in the transitive variant. These typically had *fles* 'bottle' as theme lemma, though not necessarily (142b).

- 8 prepositional instances with theme lemma *portemonnee* 'wallet' or *portefeuille* 'wallet' that express the figurative meaning 'to pay', such as (143).
- 1 prepositional instance of a building grabbing towards the air, viz. (144).

130. *Een lijnbus sloeg rechtsaf en greep de fiets.*
  a  line_bus tuned right and grabbed the bike
  'a scheduled service bus turned right and grabbed the bike.'
      (WS-U-E-A-0000063027.p.1.s.3)

131. *Het scheelde niet veel of het vuur had ook een nabijgelegen*
  it  differed not much or the fire had also a  neighboring
  *woning gegrepen.*
  house grabbed
  'It was close; the fire had nearly spread to an neighboring house.'
      (WR-P-P-G-0000047838.p.2.s.5)

132. a. *Maar zoals elke dag al  greep het peloton hen kort*
    but  like every day already grabbed the peloton them shortly
    *voor de  streep.*
    before the line
    'But, like already every day, the peloton caught up with them shortly
    before the finish line.'
       (WR-P-P-G-0000192618.p.3.s.3)

  b. *Ik knuffel ze  ook elke  dag. Elke dag grijp ik ze, zoals ik*
    I  hug  them also every day every day grab I  them like I
    *dat noem. En dat fysieke contact vind ik ontzettend*
    that call  and that fysical  contact find I  immensely
    *belangrijk.*
    important
    'I also hug them every dat. Every day, I grab them, like I call it. And I
    consider that physical contact to be immensely important.'
       (WR-P-P-G-0000104765.p.4.s.11)

c.  *De politie greep    Gert van der G.(37) nadat hij zich,      ondanks*
    the police grabbed Gert van der G.(37) after   he himself despite
    *een door de rechter opgelegd    bezoekverbod,in de tuin    van*
    a     by   the judge    imposed  visit_prohibition in the garden of
    *de voormalige zangeres(53) had  opgehouden.*
    the former      singer   (53) had  stayed
    'The police grabbed Gert van der G. (37), after he had entered the
    garden of the former singer (53), in spite of a prohibition to visiting
    restraint that had been imposed by the judge.'
        (WR-P-P-G-0000105540.p.2.s.1)

d.  *Opeens    rukte    Tuare zich     los, greep    zijn broers   en*
    suddenly  pulled  Tuare himself  loose grabbed his   brothers and
    *zijn vrouw en    kinderen  en    rende naar de junglebrug.*
    his  wife   and  children  and  ran   to   the jungle_brigde
    'Suddenly, Tuare pulled himself loose, grabbed his brothers and wife
    and children and ran to the jungle bridge.'
        (WR-P-P-B-0000000178.p.887.s.1)

e.  *Leuk voor cultuur-archeologen    die  de geest   van de jaren*
    fun  for  culture-archaeologists who the spirit   of   the years
    *zeventig proberen  te  grijpen,...*
    seventy  try        to  grab
    'Fun for culture-archaeologists that try to understand the spirit of the
    seventies.'
        (WR-P-P-G-0000061841.p.14.s.8)

f.  *Met Sneijder     erin      voor Van Damme greep    Ajax*
    with Schneijder therein  for   Van Damme grabbed Ajax
    *opnieuw     het initiatief.*
    once_again  the initiative
    'With Schneijder on the field instead of Van Damme, Ajax retook the
    initiative.'
        (WR-P-P-G-0000023279.p.8.s.2)

133. *...; en  Paulus hen ziende,dankte God en    greep    moed.*
      and  Paulus them seeing thanked God and  grabbed courage
    '...; and Paulus, seeing them, thanked God and took courage.'
        (WR-P-P-B-0000000258.p.2089.s.1)

134. *Maar  zelfs  hier  grijpen  de  jongens  het woord om de hoofddoek   te*
     but    even here grab    the boys     the word   to the head_scarf to
     *verdedigen.*
     defend
     'But even here, the boys took the floor the defend the head scarf.'
                     (WR-P-E-G-0000005520.p.109.s.1)

135. *Net  toen   de schilder haar wilde    insmeren, greep    ze opeens     een*
     just when  the painter  her   wanted rub_in     grabbed she suddenly  a
     *handvol en   gooide die   in zijn gezicht.*
     handful and  threw  that  in his  face
     'Just when the painter wanted to rub her in, she suddenly grabbed a handful
     and threw it in his face.'
                     (WR-P-P-B-0000000113.p.3213.s.1)

136. a.  *Telkens    weer   een  reden  voor Vlietland om de kijker  te grijpen*
         each_time again  a      reason for  Vlietland  to the viewer to  grab
         *en    wat     wetenswaardigheden   te spuien.*
         and  some   pieces_of_information  to spout
         'Each time another reason for Vlietland to captivate the viewer and
         spout some pieces of information.'
                     (WR-P-P-G-0000021026.p.2.s.3)

     b.  *Dit     zijn volgens       mij getuigenissen die   naar het hart*
         these  are  according_to me  testimonies   that  to    the hart grab
         *grijpen (...)*
         grab
         'These are, according to me, testimonies that speak to the heart.'
                     (WR-P-P-H-0000086843.p.16.s.2)

137. a.  *Als  er    maar één  kans   op genezing bestaat, dan grijpen we*
         if    there only  one  chance on recovery exists   then grab     we
         *die.*
         that
         'If there exists only a single chance of recovery, then we grab it.'
                     (WR-P-E-G-0000000049.p.301.s.1)

     b.  *Dat ze    iedere gelegenheid moeten grijpen.*
         that they every  opportunity  must    grab
         'That they should grab every opportunity.'
                     (WR-P-E-G-0000010103.p.143.s.1)

c. *Joden speelden een belangrijke rol (...); als onderduikers*
Jews played an important role as persons_in_hiding
*grepen zij meer dan eens naar deze mogelijkheid, zich*
grabbed they more than once to this opportunity themselves
*verdienstelijk te maken.*
useful to make.
'Jews played an important role (...); being in hiding, they often grabbed this opportunity to make themselves useful.'
      (WR-P-P-B-0000000063.p.1208.s.5)

138. a. *Het dier was kwaad geworden door de plagerijen en greep*
the animal was made become by the teasings and grabbed
*de dronken man met zijn slurf.*
the drunken man with its trunk
'The animal had been made angry by the teasing, and grabbed the drunken man with its drunk.'
      (WS-U-E-A-0000050688.p.1.s.3)

b. *Als hij met iemand samen is, grijpt die tweede naar hem*
when he with someone together is grabbed the second to him
*en hij is hulpeloos aan hem overgeleverd.*
and he is helplessly on him handed_over
'When he is together with someone, that second person grabs him and he is helplessly at his mercy.'
      (WR-P-P-B-0000000214.p.322.s.2)

c. *De vijand grijpt naar ons, verlos ons, o Heer!*
the enemy grabs to us release us o Lord
'The enemy grabs us, save us, o Lord!'
      (WR-P-P-B-0000000220.p.1111.s.1)

d. *Het meisje liep op nieuwjaarsdag in avondjurk op het*
the girl ran on New_Year's_Day in evening_dress on the
*strand te bellen, toen de twee haar graaiden.*
beach to call when the two her grasped
'the girl was on the beach on New Year's Day, calling on the phone, when the two grasped her.'
      (WR-P-P-G-0000214628.p.3.s.2)

139. a. *Na enkele meters greep een hand hem bij de rechterschouder.*
after some meters grabbed a hand him by the right_shoulder
'After a couple of meters, a hand grabbed him by the right shoulder.'
      (WR-P-P-K-0000000052.p.300.s.10)

b.  *Hij greep   meteen    naar mijn keel en   moest geld  hebben.*
    he  grabbed immediately to   my  throat and  had_to money have
    'He immediately grabbed me by the throat and wanted money.'
                (WR-P-P-G-0000658644.p.4.s.1)

c.  *(…) dat  diezelfde Wegmann hem als revanche nog even  bij*
        that the_same Wegmann  him  as revenge   still shortly with
    *de trui dacht    te graaien.*
    the shirt thought to  grobe
    '(…) that the same Wegmann wanted to grope him by the sweater, as
    an act of revenge.'
                (WR-P-P-G-0000651114.p.4.s.1)

d.  *Heckmair grabbelt Vorg bij  de broek en    redt   hem het leven.*
    Heckmair  grasps   Vorg with the pants and  saves  him  the life
    'Heckmair grasps Vorg's pants and saves his life.'
                (WR-P-P-H-0000004576.p.20.s.5)

e.  *Vrienden die  zijn  baasje te dicht naderen,  hapt  hij naar de*
    friends    that his   master too closely approach snaps he  to   the
    *vingers.*
    fingers
    'He snaps the fingers of friends that approach his master too closely'
                (WR-P-P-G-0000474959.p.2.s.3)

f.  *Ik wil ze   niet tegen   de schenen schoppen.*
    I  want them not  against  the shins     kick
    'I do not want to tread on their toes.'
                (WR-P-P-G-0000222484.p.7.s.4)

140. a.  *Hij greep   direct         naar zijn wapen, maar zijn holster was*
         he  grabbed immediately  to   his  weapon but  his holster was
         *leeg.*
         empty
         'He immediately grabbed for his gun, but his holster was empty.'
                (WR-P-P-B-0000000229.p.2113.s.2)

b.  *Doelman en   veteraan Patrick Nys, (…), graaide  tevergeefs naar de*
    keeper   and  veteran Patrick Nys            groped vainly       to   the
    *bal.*
    ball
    'Keeper and veteran Patrick Nys (…) reached for the ball in vain.'
                (WR-P-P-G-0000209555.p.2.s.4)

c. *Zeven minuten nadat (...) grabbelde de Spaanse doelman tevergeefs*
seven minutes after        grasped    the Spanish keeper  vainly
*naar een ongevaarlijk afstandsschot.*
to    a    innocuous    distance_shot
 'Seven minutes after (...), the Spanish keeper grasped vainly to an
innocuous long distance shot.'

(WR-P-P-G-0000234161.p.4.s.2)

d. *Een grauwende snuit hapte   naar hem, maar hij sprong  net*
a    growling    snout snapped at  him but    he jumped just
*op tijd  opzij.*
on time aside
'A growling snout snapped at him, but he jumped aside just in time.'

(WR-P-P-B-0000000026.p.302.s.6)

141. a. *Driekwart        van de aankopers grijpt steevast   naar bekende*
three_quarters of    the purchasers grabs invariably to    known
*merken.*
brands
'Three quarters of the purchasers invariably chooses for known
brands.'

(WR-P-P-H-0000036630.p.7.s.5)

 b. *Velen vrezen dat  Milosevic dan  naar militaire middelen  zal*
Many fear    that Milosevic then to    military means    will
*grijpen.*
grab
'Many fear that Milosevic will then turn to military means.'

(WS-U-E-A-0000010831.p.24.s.1)

c. *We grijpen dus    het beste  alternatief: de Franse concurrent*
we grab    hence the beste    alternative the French competitot
*Saint-Gobain (153.8  EUR).*
Saint-Gobain (153.8 EUR)
'So we turn to the best alternatie: the French competitor Saint-Gobain
(153.8 EUR)'

(WR-P-P-H-0000108162.p.2.s.5)

    d.   *Ik grijp   nu  zomaar voor de hand liggende eigen ervaringen;*
              I   grabbed now just     for the hand lying    own    experiences
              *"eet niet te veel   chocolade nadat je   melktanden gevallen zijn"*
              eat  not  too much chocolate after   your milk_teeth fallen   are
              'I'm just picking obvious personal experiences; don't eat too much
              chocolate, after your milk teeth have fallen out.'
                        (WR-P-P-H-0000050373.p.12.s.4)

142. a.   *Iedereen grijpt wel eens naar de   fles.*
           everyone grabs  PART once to    the   bottle
           'Everyone occasionally reaches for the bottle.'
                      (WR-P-E-G-0000001985.p.189.s.1)

     b.   *De hoge officier die… Hij greep   ook regelmatig   naar het glas,*
           the high officer   that he grabbed also regularly    to    the glass
           *maar daar   maakte hij geen geheim van.*
           but  there  made   he no    secret  of
           'The high officer that… He also regularly reached for the glass, but he
           made no secret out of it.'
                      (WR-P-E-G-0000006809.p.90.s.1)

143. *Hoe blijven de verschillende ngo's   erin    slagen  om mensen*
     how  stay   the various        NGO's therein succeed to people
     *telkens     opnieuw naar hun  portemonnee te doen   grijpen?*
     each_time again   to    their wallet         to  do    grab
     'How do the various NGO's succeed time and again to make people reach
     for their wallet?'
                      (WR-P-P-G-0000216228.p.6.s.1)

144. *Griezelig   hoog zijn ze,  de twee rollercoastertorens  die  op de*
     Gruesomely high are   they, the two  rollercoaster_towers that  on the
     *CSM-terreinen naar de Limburgse   lucht grijpen.*
     CSM-terrains   to    the Limburgian air   grab
       'Gruesomely high they are, the two rollercoaster towers that grab into
       the Limburgian air on the CSM-terrains.'
                      (WR-P-P-H-0000073550.p.3.s.1)

Finally, 10 instances of idiomatic expressions of the verb *happen* 'snap', viz. of quasi-noun-incorporation that referred to a cultural activity (Booij 2009), as in (145), were also removed, as well as the following (semi-)idiomatic uses of the verb *schoppen* 'kick'.

- 9 instances of *met kind geschopt*, meaning 'made pregnant', as in (146).
- As with *happen* 'snap', 1 transitive instance of quasi-noun-incorporation that referred to a cultural activity as in (147) (Booij 2009).
- 1 instance of *een schop schoppen* 'to kick a kick' (148).
- 1847 transitive instances of *het schoppen* with the figurative meaning manage to become or to get', as in (149).
- 373 transitive instances of *herrie, keet, branie, problemen, schandaal, een scène, rel,… schoppen* 'cause trouble', as in (150).
- 13 psrepositional instances with the figurative meaning 'criticize' in (151).

145. a.  *Overal      waar  het meisje kwam, zag  ze  hen  zaklopen     of*
         everywhere where the girl    came  saw she them sack_running or
         *koek happen.*
         cake snapping
         'Everywhere the girl came, she saw them running a sack race or
         playing bite-the-cake.'
                      (WR-P-P-H-0000175600.p.2.s.5)

     c.  *We gaan altijd  met het hele    bedrijf    haring happen (…)*
         we  go   always with the entire company herring snap
         'We always go out with the entire company to eat herring.'
                      (WR-P-P-G-0000081989.p.11.s.3)

146. *Uiteindelijk kan  het toch niemand iets       schelen dat  Freud zijn*
     ultimately   can it still no_one  something differ   that Freud his
     *schoonzus     met kind heeft geschopt?'*
     sister_in_law with child has   kicked
     'In the end, who cares that Freud made his sister in law pregnant?'
                      (WR-P-P-B-0000000153.p.1630.s.3)

147. *En   voor de gelegenheid vonden de vrienden  een  nieuwe*
     and for the occasion    invented the friends  a    new
     *volkssport    uit: slesjskippen, of pantoffel schoppen.*
     national_sport PART slesjskippen or slipper   kicking
                      (WR-P-P-G-0000415997.p.2.s.2)

148. *Schotten die  hoekschoppen schoppen, zijn Hoekschotten.*
     Scotsmen that corners        kick      are  Corner_scotsmen
     'Scotsmen that take corners, are Corner-Scotsmen.'
                      (WR-P-P-H-0000107645.p.6.s.4)

149.  a.   *Hailey heeft het tot schrijver  van  CSI-New York geschopt.*
             Hailey has it  to  writer     of    CSI-New York kicked
             'Hailey has made it to writer of CSI-New York.'
                          (WR-P-P-G-0000221387.p.6.s.2)

    d.   *Max merkt  al       gauw dat Stan bokstalent    heeft en   dat*
          Max notices already soon   that Stan boxing_talent has  and  that
          *hij het ver  kan  schoppen.*
          he it  far   can   kick
          'Max quickly notices that Stan had boxing talent and might go far.'
                          (WR-P-P-G-0000227997.p.2.s.2)

150.  *De politie hield    ook twaalf heethoofden  aan  die   keet    schopten*
      the police arrested also twelve hotheads      part that racket kicked
      *in de buurt        van het stadion.*
      in the neighborhood of    the stadion
      'The police also arrested twelve hotheads that were kicking up a racket in
      the neighborhood of the stadion.'
                          (WR-P-P-G-0000222825.p.13.s.1)

151.  *Van Mulukom  schopte ook  naar de organisatoren van  de grotere rondes.*
      Van Mulukom kicked  also to   the organizers    of    the big   tours.
      'Van Mulukom also criticized the organizers of the important tours.'
                          (WR-P-P-G-0000655022.p.3.s.1)

### 5.2.5.2      Borderline retained instances

The following instances were retained in the dataset after some consideration.

- Instances where it is clear from the context that the act of grabbing was successful, as these regularly appeared in the transitive and prepositional variant , as in (152a-b). Such instances were also retained for *graaien* 'grasp', as in (152c-d), *grabbelen* 'scramble', as in (152e-f), *happen* 'snap', as in (152g-h), and schoppen 'kick', as in (152i-j).
- Instances meaning 'to grab power' (153a-b) and 'to grab victory' (153c-d) were retained, as these regularly occurred both in the transitive and prepositional variant. One such instance was also retained for the near-synonymous verb *graaien* 'grasp', viz. (152e).

- Instances of *(naar) de wapens, (naar) de pen grijpen, (naar) de telefoon grijpen* in the figurative uses of respectively 'to take up arms', 'to start writing' and 'to start phoning' were retained, because these regularly appeared in the transitive and prepositional variant (154).

152. a. *De 49-jarige    Bruno Kestens greep    een mes en    stak    er*
     the 49_year_old Bruno Kestens grabbed a    knife and stabbed there
     *zijn 40-jarige    echtgenote mee in de hals.*
     his   40_year_old wife       with in the neck
     'The 49 year old Bruno Kestens grabbed a knife and stabbed his 40 year old wife in the neck with it.'
     (WR-P-P-G-0000508873.p.2.s.2)

   b. *Daarop   greep   de dader      naar een mes en    stak    het*
      thereupon grabbed the perpetrator to  a    knife and stabbed it
      *in Frederik's buik.*
      in Frederik's belly
      'Thereupon, the perpetrator grabbed a knife and stabbed it into Frederik's belly.'
      (WR-P-P-G-0000359660.p.4.s.5)

   c. *'Diefstal is   vooral een zaak     van zakkenrollers, of mensen die*
      theft   is mostly a    business of   pickpockets    or people that
      *mobiele telefoons   graaien', zegt Johnny.*
      mobile  phones     grobe    says Johnny
      '"Theft is mostly a business for pickpockets, or people that grope mobile phones", says Johnny.'
      (WR-P-P-G-0000045456.p.4.s.1)

   d. *Dan graait Robbert naar zijn sigaretten en   houdt het pakje*
      then grasps Robbert to   his  cigarettes and holds the package
      *voor   mijn neus.*
      before my  nose
      'Then Robbert gropes his cigarettes and holds the package in front of my nose.'
      (WR-P-P-B-0000000224.p.81.s.2)

   e. *Ze grabbelden    geld en   sloegen op de vlucht.*
      they scrambled   money and hit      on the flight
      'They grasped money and took to flight.'
      (WR-P-P-H-0000102713.p.3.s.6)

f.  *Ze grabbelt naar de muntstukken. Stopt ze   in de plastic pot.*
    she scrambles to the coin_pieces   puts   them in the plastic pot
    'She grasps the pieces of coin. Puts them in the plastic pot.'
    (WR-P-P-H-0000020589.p.17.s.3)

g.  *Hij hapte   opgelucht adem na   een zege   zonder glans.*
    he snapped relieved    breath after a     victory without luster
    'he gasped for breath in relieve, after a lackluster victory.'
    (WR-P-P-G-0000523686.p.3.s.2)

h.  *Ik kwam tot bezinning, hapte   naar lucht, propte de brief in mijn*
    I   came to sense     snapped to     air     stuffed the letter in my
    *zak, (...)*
    pocket
    'I came to my senses, gasped for air, stuffed the letter in my pocket.'
    (WR-P-P-B-0000000007.p.2596.s.14)

i.  *Hij  sloeg  de man neer   en   schopte hem nog enkele keren.*
    he   hit    the man down and   kicked  him still some  times
    'He struck the man down and then kicked him a number of times.'
    (WR-P-P-G-0000386776.p.2.s.4)

j.  *Ze   bleven me maar slaanen    stampen met    hun zware*
    they stayed me just  hit  and  stamp  with   their heavy
    *bottines. Niemand schopt zelfs zo naar een  beest, denk  ik.*
    high-lows. No_one   kicks  even so  to  a    beast  think I.
    'They just kept on hitting and stomping me with their heavy high-
    lows. No one even kicks a beast like that, I think.'
    (WR-P-P-G-0000598434.p.6.s.5)

153. a.  *Grijpen de vrouwen de macht op de VRT-nieuwsdienst?*
         grab     the women  the power on the VRT-news_service
         'Do the women grabbed the power on the VRT news service?'
         (WR-P-P-G-0000668510.p.7.s.2)

     b.  *Vrouwen die  een hardere lijn willen, grijpen bij  de groene*
         women   that a   harder  line want    grab    with the green
         *partij naar de macht.*
         party  to   the power
         'Women that want a harder line, grab the power in the green party.'
         (WR-P-P-G-0000395793.p.7.s.3)

c. *Eerst was er      Oscar Freire, klaar  om zijn  derde  ritoverwinning*
first was there Oscar Freire ready  to his   third   stage_victory
*te  grijpen, (...)*
to  grab
'First there was Oscar Freire, ready to grab his third stage victory.'
          (WR-P-P-G-0000646010.p.4.s.1)

d. *Na  een  spannende  en    ultralange dubbel grepen  Melanie en*
after a   suspenseful and   ultra_long double grabbed Melanie  and
*Rosaline   naar het goud na   een  zege  in  drie sets.*
Rosaline   to   the gold after a    victory in three sets
'After a suspenseful and ultra-long double, Melanie and Rosaline
grabbed the gold after a victory in three sets.'
          (WR-P-P-G-0000404835.p.2.s.5)

e. *De La Fuente graait  de  premie Jacques Goddet op  de  top van  de*
De La Fuente gropes the bonus  Jacques Goddet on  the top of    the
*Tourmalet.*
Tourmalet
'De La Fuente grabs the bonus Jacquies Goddet on the top of the
Tourmalet.'
          (WR-P-P-G-0000644903.p.3.s.2)

154. a. *(...) "als ze   in een  ander  land zouden  wonen,  ze*
              if    they in an   other  land would   live      they
*allang          de wapens    hadden gegrepen".*
already_long   the weapons    had     grabbed
'If they had lived in aother country, they would have already taken up
arms a long time ago.'
          (WR-P-P-H-0000148082.p.13.s.5)

b. *Deze operatie,  de eerste keer dat   nieuwe  bekeerden naar de*
this operation the first   time that new      converts   to    the
*wapens hadden  gegrepen,    (...)*
weapons had      grabbed
'This operation, the first that the new converts had taken up arms, (...)'
          (WR-P-P-B-0000000065.p.1871.s.9)

c. *Zij  grepen  de  pen  en   publiceerden hun verhalen.*
they grabbed their pen  and  published     their stories
'They took up their pens and published their stories.'
          (WR-P-P-G-0000134183.p.8.s.2)

d.	*Ik greep	weer	naar de pen en	schreef	mijn gedachten op*
I	grabbed again	to	the pen and	wrote	my	thoughts	up
*in  een  artikel...*
in  an   article
'I took up my pen again and wrote my thoughts down in an article.'
(WR-P-P-B-0000000205.p.1380.s.1)

e.	*Toen ik terugkwam in het Varkenshok greep	ik de telefoon*
when I   returned   in the Varkenshok grabbed I   the phone
*en   boekte  haastig  een  enkele  reis naar Alaska.*
and booked hastily  a   one-way trip to   Alaska
'When I had returned to the Varkenshok, I grabbed the telephone and
hastily booked a one-way trip to Alaska.'
(WR-P-P-B-0000000210.p.76.s.1)

f.	*Hij greep	naar de telefoon, vroeg  de telefoniste	de*
he grabbed to   the phone   asked the telephone_operator the
*Belgische koning in Laeken te  bellen en   viel  nog  geen minuut*
Belgian   king  in Laeken to call   and fell  still no   minute
*later onmiddellijk  met de deur in huis.*
later immediately  with the door in  house
'He grabbed the telephone, asked the phone operator to call the
Belgian king in Laeken and came straight to the point, not a minute
later.'
(WR-P-P-G-0000095220.p.4.s.5)

## 5.2.6    Telephonic verbs

The telephonic verbs are the synonyms *bellen* 'phone', *telefoneren* 'phone' and *opbellen* 'phone'. *Opbellen* 'phone' may be the odd verb out, since it is a separable particle verb. In fact, it could be argued that this verb should not be analyzed as a separate verb at all, but rather as a resultative usage of the verb *bellen* 'phone', meaning something akin to 'phoning someone, causing them to rise up'. In that case, the verb should not occur in the prepositional variant. It did, however, as in (155). As such, we considered it a verb in its own right, as does the Alpino-parser (van Noord 2006) and most dictionaries (e.g. den Boon & Geeraerts 2005; Weijnen & Ficq-Weijnen 2008). Still, its status as a separable particle verb is a good reason to be weary of it behaving like an outlier in our analyses. We come back to this in Section 6.1.

155. *U    belt    naar hen op:   U    hoort  nu   in een reconstructie (...)*
     you  phone to   them PART you  hear   now  in a     reconstruction
     'You phone them: you now hear in a reconstruction (...)'

### 5.2.6.1      Removed Instances

Prepositional instances that involve phoning to a network, an operator or e.g. 'from a land line to mobile', such as (156), were removed for all telephonic verbs. These numbered 16 in total. One such instance of the verb *bellen* 'phone' did exhibit the transitive variant, viz. (156e), yet this concerned an isolated infinitive in a headline, so it was considered exceptional. It was also removed for consistency. For the same reason, the single such instance of *opbellen* 'phone', viz. (156f), which also exhibited the transitive variant, was also removed.

156. a.   *Voor 18,8 cent per minuut bel    je    op eender welk   moment*
          for   18.8 cents per minute  phone you  on any     which moment
          *naar eender welk   netwerk – vast     of  mobiel.*
          to   any     which network   landline or  mobile
          'For 18.8 cents per minute, you call on any moment to any network –
          landline or mobile.'
                        (WR-P-P-G-0000221050.p.1.s.3)

     b.   *Bovendien zijn er   steeds meer operatoren die het  'any time,*
          Moreover   are there ever   more operators   that the  any time,
          *any network'-principe hanteren, waarbij hetzelfde tarief  geldt*
          any network principle maintain  whereby the_same tariff   holds
          *wanneer je    naar een andere   operator belt.*
          when    you  to   an  other     operator calls
          'Moreover, there are ever more operators that maintain the 'any time,
          any network'-principle, whereby the same tariff applies when you call
          to a different operator.'
                        (WR-P-P-G-0000221050.p.1.s.3)

     c.   *LDV: Maar als je    nu   van mobiel naar mobiel belt?*
          LDV  but  if  you  now  of  mobile to   mobile phone
          'LDV: but what if you phone from mobile to mobile?'
                        (WS-U-E-A-0000235514.p.2.s.1)

d. *(...)Parlino, een internetdienst    waarmee    je (...) voor een*
   Parlino an internet_service wherewith you    for a
   *gunstig    tarief    naar vaste of mobiele lijnen kunt telefoneren.*
   favorable fee    to    land or mobile lines can phone
   '(...) Parlino, an internet service with which you (...) can phone to
   landlines or mobile numbers for a favorable fee.'
         (WR-P-P-G-0000185003.p.9.s.1)

e. *Ander netwerk bellen.*
   other   network phone
   'Phone another network.'
         (WS-U-E-A-0000245285.head.1.s.1)

f. *Een ISP   opbellen om via zijn   netwerk Internet-toegang te*
   an ISP phone    to via his   network internet_access    to
   *krijgen (...)*
   get
   'Calling an ISP to get internet access via its network (...)'
         (WR-P-P-H-0000037613.p.2.s.4)


220 prepositional instances of *naar huis bellen/telefoneren* 'phone home', *naar hier/daar/ginder bellen/telefoneren* 'phone to here/there/yonder', as in (157), were removed for *bellen* 'phone' and *telefoneren* 'phone'. These did not occur for *opbellen* 'phone'.


157. a. *Dan kan ik naar huis    bellen.*
    then can I   to    house phone
    'Then I can phone home.'
          (WR-P-E-G-0000002332.p.109.s.1)

b. *Ik zal   Bianca zeggen dat   ze niet meer naar hier moet bellen.*
   I will Bianca say   that she not more to   here must call
   'I'll tell Bianca that she shouldn't phone to here anymore.'
          (WR-P-E-G-0000001976.p.291.s.1)

c. *Van Mierlo heeft naar daar   gebeld.*
   Van Mierlo has to   there   phoned
   'Van Mierlo has phoned to there.'
          (WR-P-E-G-0000002167.p.666.s.1)

d.  *Soms       telefoneren ze    naar huis.*
    sometimes  phone        they  to   house
    'Sometimes, they phone home.'
                (WR-P-P-H-0000050074.p.7.s.11)

e.  *Ik wou     vragen of    je   naar ginder kon     telefoneren.*
    I   wanted  ask   whether you  to   there  could   phone
    'I wanted to ask whether you can phone to there.'
                (WR-P-E-G-0000000709.p.138.s.2)

The following instances were only removed for *bellen* 'phone', since they did not occur for the other two verbs.

- 4 transitive instances of the commercial slogan *even Apeldoorn bellen* 'just call Apeldoorn' of the insurance firm *Centraal Beheer* (158a). Four instances of phoning to Apeldoorn, where the slogan is not retained verbatim, however, such as (158b-c), were retained, as one of them did exhibit the prepositional variant, viz. (158c).
- 324 transitive instances of ordering an inanimate theme, such as (159).

158. a.  *Even Apeldoorn       bellen,  werkt  niet   altijd.*
         shortly Apeldoorn      phone    works  not    always
         'Quickly calling Apeldoorn doesn't always work.'
                    (WR-P-P-G-0000066860.p.8.s.2)

    b.  *Zowat alles        werd een tiental    keer   nagerekend waarbij*
        almost everything   was  an  about_ten  times  recalculated whereby
        *de begrotingsminister, (...)   meer  dan  eens "Apeldoorn moest*
        the Minister_of_Finance, (...)  more  than once Apeldoorn  had_to
        *bellen"(liberaal jargon voor kabinetschef  John Combrez)*
        call"  liberal   jargon  for   chief_of_staff John Combrez
        'Almost everything was checked about ten times, whereby the Minister of Finances had to "call Apeldoorn" multiple times (liberal jargon for chief of staff John Combrez).'
                    (WR-P-P-G-0000245067.p.9.s.4)

c. *En   volgens        de betrokken verzekeringsmaatschappij wordt*
and  according_to the involved   insurance_company        is
*er    ook  nog  altijd  ruimschoots naar Apeldoorn gebeld.*
there also  still  always amply        to    Apeldoorn  called
'And according to the involved insurance company, Apeldoorn is still often being called.'
                    (WR-P-P-G-0000078873.p.12.s.1)

159. *Ik  kan  ook  een  taxi  bellen.*
     I    can  also  a      taxi  call
     'I can also call a taxi?'
                    (WR-P-E-G-0000000900.p.621.s.1)

### 5.2.6.2     Borderline retained instances

The following instances were retained in the dataset after some consideration.

- Instances of calling an animate or collective theme to come over, as these appeared in both variants, such as (160).
- Instances where the theme is a place, except for *naar huis* 'home', *naar hier* 'to here' and *naar daar* 'to there', as these appeared in both variants, as in (161). This includes instances of the theme lemma *huis* 'house' with a determiner, as these also did appear in both variants (161c-d).
- Instances such as (162). At first sight, these look like ditransitive rather than transitive instances. However, since they also appear in the prepositional instances as in (162b), these instances were nonetheless retained.
- Instances where the theme is a telephone apparatus, as they appear in both variants (163).
- Instances where the theme was being phoned *op zijn/haar gsm, vaste lijn, kantoor, werk,...* 'on his/her cell phone, land line, office, job,...', as in (164).

160. a. *Als  ik  jullie hier  nog  eens zie, bel    ik  de politie.*
        if    I    you   here  still  once see phone I    the police
        'If I see you here one more time, I'll call the police.'
                    (WR-P-E-G-0000000354.p.101.s.1)

b.  *Terwijl (...) belden  de buren     vanwege   de overlast  naar*
    While       phoned  the neighbors because_of the nuisance  to
    *de politie. Die drong             prompt    de woning binnen en*
    the police  that  entered_by_force promptly  the house  PART and
    *schakelde de  computer uit.*
    turned     the computer off
    'While (...), the neighbors called the police because of the nuisance.
    They promptly forced their ways into the house and turned off the
    computer.'
          (WR-P-P-G-0000192721.p.3.s.1)

c.  *Ze (...) vraagt of       ze de politie mag telefoneren. Een half*
    she      asks  whether she the police may phone       a      half
    *uur    later    komt   een    patrouille      ter      plaatse.*
    hour   later    comes  a      patrol     to_the place
    'She (...) asks whether she can call the police. Half an hour later, a
    patrol arrives at the scene.'
          (WR-P-P-G-0000205773.p.3.s.4)

d.  *Pas   na    enige  tijd        was hij genoeg bekomen van  de*
    only after some   time        was he  enough recovered of    the
    *emoties   om naar de politie te  telefoneren. Amper   een  paar*
    emotions  to to    the police to phone        just    a    few
    *minuten later was die   al       ter     plaatse.*
    minutes later was  that  already  to_the  place
    'Only after some time time, he had sufficiently recovered from the
    emotions to call the police. Only a few hours later, they were at the
    scene.'
          (WR-P-P-G-0000653765.p.6.s.5)

e.  *Een getuige belde    de politie op  en   die kon   de man*
    a   witness phoned  the police part and  that could the man
    *onderscheppen.*
    intercept
    'A witness phoned the police and they could intercept the man.'
          (WR-P-P-G-0000645961.p.2.s.4)

161. a.  *Hij ging in de werkkamer van  zijn vriend zitten om Irak te*
         he  went in the study       of   his  friend sit    to Irak to
         *bellen.*
         phone
         'He went into the study of his friend and sat down to call Irak.'
               (WR-P-P-B-0000000203.p.1614.s.1)

b. *De joodse familie probeert naar New York te  bellen (...)*
the jewish family tries     to    New York to phone.'
'The jewish family tries to phone to New York.'
              (WR-P-E-G-0000000309.p.93.s.1)

c. *Ik belde   het huis   van haar ouders, maar ze was er    niet.*
I   phoned the house of   her parent  but  she was there not
'I phoned the house of her parents, but she wasn't there.'
              (WR-P-P-B-0000000149.p.846.s.8)

d. *(...)dat  de Bosnische leider Radovan Karadzic de fout     maakte*
    that the Bosnian   leader Radovan Karadzic the mistake  made
*naar het huis   van de familie Milosevic te  bellen (...)*
to    the house of   the family Milosevic to phone
' (...) that the Bosnian leader Radovan Karadzic made the mistake of
calling to the house of the family Milosevic. (...)'
              (WR-P-P-G-0000179297.p.13.s.1)

e. *Elke   dag telefoneer ik naar zijn vakantieadres.*
every day phone     I  to   his  vacation_address
'Every day, I phone his vacation address.'
              (WR-P-P-G-0000610366.p.17.s.8)

f. *De secretaris-generaal van de Verenigde Naties    belde    Damascus*
the Secretary-general  of   the United     Nations phoned Damascus
*op (...)*
PART
'The secretary-general of the United Nations phoned Damascus (...)'
              (WR-P-P-H-0000125263.p.4.s.4)

162. a. *Even later belde    mijn uitgever me, dat de distributie*
shortly later phoned  my   publisher me  that the distribution
*vertraging had  opgelopen.*
delay      had  run_into
'Shortly later, my publisher called me that the distribution had run into
delays.'
              (WR-P-P-G-0000181698.p.7.s.5)

b. *Een bewoonster van de Reeboklaan in Kapellen belde*
a resident of the Reeboklaan in Kapellen phoned
*maandagmiddag naar de politie, dat ze beneden inbrekers*
mondaymorning to the police that she downstairs burglars
*hoorde.*
heared
'A resident of the Reeboklaan in Kapellen called the police yesterday, that she heared burglars downstairs.'
   (WR-P-P-G-0000289589.p.2.s.1)

c. *Maar toen ik haar het nieuws telefoneerde, werd ze gek.*
but when I her the news phoned become she insane
'But when I phoned her the news, she freaked out.'
   (WR-P-P-G-0000399712.p.2.s.2)

163. a. *Moet ik z'n gsm bellen?*
must I his mobile_phone phone
'Should I call his mobile phone?'
   (WR-P-E-G-0000000900.p.365.s.1)

b. *Bel naar haar gsm.*
phone to her mobile_phone
'Call her mobile phone.'
   (WR-P-E-G-0000001016.p.389.s.1)

c. *Het slachtoffer telefoneerde immers naar haar eigen GSM (...)*
the victim phoned after_all to her own mobile
'For the victim phoned her own mobile phone (...)'
   (WR-P-P-G-0000277444.p.2.s.5)

164. *U kunt uw broer bellen op z'n werk of z'n mobiel.*
you can your brother phone on his work or his cell_phone
'You can call your brother at his work place or on his cell phone.'
   (WR-P-P-G-0000110267.p.8.s.4)

## 5.2.7   *Jagen* 'hunt'

### 5.2.7.1   Removed instances

The following instances, which are all figurative and/or idiomatic uses which do not allow the alternative variant, were removed for *jagen* 'hunt'.

- 9 prepositional instances with the figurative meaning 'attempt to reach', as in (165).
- 9 transitive instances of quasi-noun-incorporation, such as (166) (Booij 2009).
- 1 transitive instance where *jagen* means 'to mentally hound', as in (167).

165. *Wie alleen naar pieken jaagt, komt van een kouwe kermis thuis.*
    Who only to peaks hunts comes of a cold fair home
    'We only aims for peaks, is up for a rude awakening.'
                (WR-P-P-H-0000064935.p.12.s.12)

166. *Na het vermoeiende koopjes jagen is slapen in een mooi*
    after the tiring bargains hunting is sleeping in a beautiful
    *hotel in Engelse stijl geen overdreven luxe.*
    hotel in English style no excessive luxury
    'After a tiring day of shopping, going to sleep in a beautiful English-style hotel is no excessive luxury.'
                (WR-P-P-G-0000187503.p.14.s.1)

167. *(...)zodat het geruis van een gedreven blad hen jagen zal, en*
       such_that the rustling of a floated leaf them hunt will and
    *zij zullen vluchten (...)*
    they will flee
    '(...) such that the rustling of a floated leaf will hound them, and they will flee (...)'
                (WR-P-P-B-0000000490.p.1692.s.1)

### 5.2.7.2      Retained instances

This only left 37 instances of the verb *jagen* 'hunt', with the majority being removed because they were used in a resultative construction.[60] A transitive and prepositional instance is shown in (168).

---

[60] *Jagen* 'hunt' is also used with the preposition *op* 'on' see Table 2.

168. a.     *Terwijl de hertog robben en ijsberen    ging jagen, deed mijn*
         while    the duke   seals   and polar_bears went hunt   did   my
         *vader aan wetenschappelijk werk, zegt Gaston de Gerlache.*
         father on    scientific        work,   says Gaston de Gerlache
         'While the duke went hunting for seals and polar bears, my father
         conducted scientific research, says Gaston de Gerlache.'
                  (WR-P-P-H-0000018295.p.11.s.3)

     b.     *Klein en behendig jagen ze   over de dunste takken    naar*
         small and agile     hunt   they over the thinnest branches   for
         *malse insecten.*
         tender insects
         'Small and agile, they hunt for tender insects on the thinnest of
         branches.'
                  (WR-P-E-G-0000002676.p.233.s.1)

## 5.2.8     *Vissen* 'fish'

### 5.2.8.1     Removed instances

The following instances were removed for *vissen* 'fish'.

- 5 transitive instances where the theme is explicitly quantified, such as (169).
- 18 transitive instances of quasi-noun-incorporation such as (170) (Booij 2009).
- 3 transitive instances where a team is trying to get an athlete to join them, as in (171).
- 41 prepositional instances where *vissen* is used figuratively and can be translated as 'try to get to know something about' or 'try to obtain', as in (172).

This concludes the manual checking of the data. Chapter 6 can now finally turn to testing the lectal and semantic hypotheses.

169.  *(...)om de komende maanden ruim    drie    miljoen kilo kokkels op*
          to the coming    months   at_least three  million kilo cockles on
      *de Waddenzee   te  vissen.*
      the Wadden_Sea to  fish
      '(...) to fish at least three million cockles on the Wadden Sea the coming
      months.'
                    (WR-P-P-G-0000143655.p.2.s.2)

170.  *Roger baatte drie kramen uit: eendjes      vissen, potten gooien      en*
      Roger ran    three stalls  PART little_duck fishing pot    throwing and
      *koordje    trekken.*
      little_cord pull
      'Roger ran three stalls: fishing for little ducks, throwing pots and pulling
      cords.'
                    (WR-P-P-G-0000510374.p.4.s.1)

171.  *De   club vist    naar David Coene, spelverdeler bij*
      The club fishes to    David Coene  playmaker   with
      *tweedeprovincialer      Ede*
      second_provincial_club  Ede
      'The club is fishing for David Coene, playmaker of Ede, a club from the
      second provincial league.'
                    (WR-P-P-G-0000557358.p.15.s.3)

172.  a.    *Niet naar mijn seksleven vissen, hé jongen?*
            not to    my  sex_life    fish   he boy
            'Don't angle for my sex life, he boy?'
                        (WR-P-E-G-0000005213.p.17.s.1)

      b.    *Shana vist    naar mijn aandacht door fors      aan haar*
            Shana fishes to    my  attention by   forcefully on  her
            *kralenketting   te  trekken (...)*
            string_of_beads to  pull
            'Shana angles for my attention by forcefully pulling her string of
            beads.'
                        (WR-P-P-G-0000516333.p.24.s.6)


### 5.2.8.2    Borderline retained instances

As with *grijpen* 'grab', and *graaien* 'grasp', *grabbelen* 'scramble', *happen* 'snap'
and *schoppen* 'kick', one instance where it is clear from the context that the act of
fishing had been successful, was retained, viz. (173).

173. *Als kind hielp Bart Desmidt al      de garnalen pellen die zijn*
     as   child helped Bart Desmidt already the shrimps   peal   that his
     *grootvader en   oom   hadden gevist.*
     grandfather and   uncle   had       fished
     'As a child, Bart Desmidt was already helping to peal the shrimps that his
     grandfather and uncle had fished.'
                    (WR-P-P-G-0000654178.p.3.s.3)

# 6    Testing the lectal and semantic hypotheses

This chapter tests the predictions based on the lectal and semantic hypotheses formulated in Chapter 4. The first section tests the predictions at the level of the preposition, the second at the level of the verb, and the third at the level of the object. The fourth section interprets the results and ends this chapter with some preliminary conclusions.[61]

## 6.1    Level of the preposition

At the level of the preposition, we will test the predictions Lec1, Lec2 and Sem1.1 (cf. Section 4.4). To do so, separate regression models were built based on the Netherlandic and Belgian data using the measure VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION (cf. Equation 1 in Subsection 4.2.2.1). This measure is not identical for the Netherlands and Belgium, as we used different collexemes for both countries.

It could then be asked whether the Netherlandic and Belgian regression models are still comparable. We argue that they are, because the procedure underlying both measures is still identical, i.e. select the top 5 collexemes and calculate the difference in average distributional proximity. Moreover, it is certainly possible that our speculative interpretations of the collexemes in Subsection 4.2.2.1 were not entirely accurate, and there is actually a slightly different semantic contrast at play in the Netherlands than in Belgium. If so, we will have anticipated this by using regiolectal-specific collexemes. Lastly, even conceding that the use of different measures reduces the comparability of both models, we see no reason why this

---

[61] For the analyses presented in this chapter, we made use of the R-packages *tidyr, car, Hmisc, lme4, effects* and all packages on which these depend (Wickham and Henry 2019; Fox and Weisberg 2011; Harrell 2017; Bates et al. 2015; Fox 2003).

would increase the likelihood of Lec1 and Lec2 to be confirmed. In other words, the use of regiolect-specific measures for VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION does not unfairly benefit the Netherlandic models compared to the Belgian models. In fact, using a unitary measure for both regiolects would run the risk of the measure being more optimized for one regiolect over the other. The same holds for the other semantic coherence measures proposed in Chapter 4.

Figure 6 shows the ranking of the verbs according to their VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION in Belgium and the Netherlands. The rankings indeed differ, if only slightly. In both rankings, *bellen* 'phone', *jagen* 'hunt' and *schoppen* 'kick' are ranked highly, followed by *telefoneren* 'phone', *opbellen* 'phone', *happen* 'snap' and *vissen* 'fish', while *grabbelen* 'scramble', *grijpen* 'grab' and *graaien* 'grasp' are ranked fairly low, and *zoeken* 'search', *verlangen* 'desire' and *peilen* 'gauge' lowest.
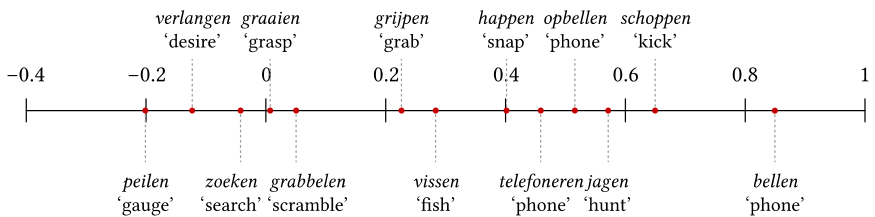


a. The Netherlandic ranking
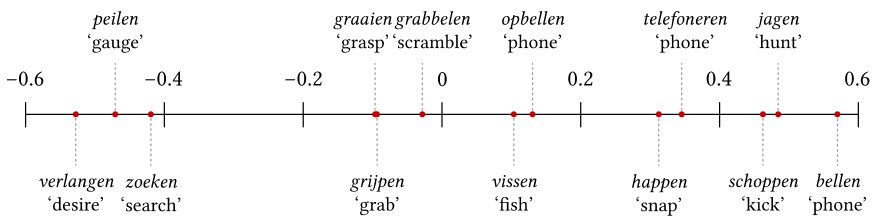


b. The Belgian ranking

Figure 6: Rankings of the verbs according to their value for VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION.

While the collostructional analyses were run separately for the Netherlands and Belgium, the distributional vectors were calculated based on the entire corpus. The underlying assumption is that the constructional meaning of both variants may vary for each regiolect, but we expect the lexical meanings of the verbs to be on average fairly stable. This is also assumed in Levshina and Heylen (2014: 29). We did also build separate distributional vectors based on the Netherlandic and Belgian subcorpora, but these did not yield qualitatively different results than the vectors based on the entire corpus.

A Netherlandic and Belgian regression model are then composed with VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION as fixed effect. As data, we use all instances that were retained as interchangeable in the previous chapter. Since all instances of the same verb have an identical value for VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION, the variable VERB is also added as a random effect with random intercepts (Speelman, Heylen and Geeraerts 2018: 2). This variable has a separate level for each verb. Furthermore, the variables THEME COMPLEXITY, VERB-THEME ORDER and an interaction between both are added as fixed effects to control for the influence of complexity (see Subsection 4.3.2.1). The categorical variable VERB-THEME ORDER distinguishes between the instances where the theme precedes the verb (*theme-verb*) and those where the verb precedes the theme (*verb-theme*). This variable is implemented through dummy coding, also known as treatment contrasts, with *theme-verb* as reference level (Gries 2013: 270–314). This means that its parameter is set to 0 for instances where the theme precedes the verb, and 1 for instances where the verb precedes the theme.

Finally, we add CORPUS COMPONENT as an additional random effect with random intercepts to control for the influence of register. It would also have been possible to consider register a fixed effect. In that case, one would typically use coarse-grained levels that are exhaustive and can hence be used in a follow-up study, e.g. *formal register* vs. *informal register*. However, we prefer to directly use the more fine-grained distinction between individual corpus components, which means that the levels are not repeatable when a follow-up study would use a different corpus. As such, we opt for a random effect. For a discussion of the merits of both approaches, see Speelman, Heylen and Geeraerts (2018: 3).

The specifications of the Netherlandic and Belgian models can be found respectively in  Table 13 and Table 14. Multicollinearity was not found to be a problem, with all Variance Inflation Factors (VIF) well below 5 (Levshina 2015: 160).[62] The C-indexes indicate acceptable discrimination, although they are rather on the low side (Hosmer and Lemeshow 2000: 162). Figure 7 shows the effect plot of VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION in the Netherlands and Belgium.

---

[62] All VIF's reported in this theses were calculated for the same models with the interaction between THEME COMPLEXITY and VERB-THEME ORDER is removed.

AIC: 23,740.9      Transitive observations: 24,346
C-index: 0.745      Prepositional observations (success level): 5,216

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | -0.97 | 0.71 | -1.37 | 0.1724 |
| VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION | | -2.03 | 1.40 | -1.45 | 0.1482 |
| THEME COMPLEXITY | | -0.12 | 0.04 | -3.32 | 0.0009 |
| VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.56 | 0.05 | 11.27 | < 0.0001 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.35 | 0.04 | 7.99 | < 0.0001 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| VERB | 13 | 5.20 | 2.28 |
| CORPUS COMPONENT | 15 | 0.03 | 0.18 |

Table 13: Specifications of a regression model at the preposition level, based on the **Netherlandic** data.

AIC: 55,703      Transitive observations: 51,792
C-index: 0.700      Prepositional observations (success level): 12,314

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | -0.91 | 0.46 | -1.96 | 0.0495 |
| VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION | | -1.00 | 0.71 | -1.42 | 0.1560 |
| THEME COMPLEXITY | | 0.01 | 0.02 | 0.29 | 0.7694 |
| VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.37 | 0.03 | 11.01 | < 0.0001 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.27 | 0.03 | 8.97 | < 0.0001 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| VERB | 13 | 0.12 | 0.34 |
| CORPUS COMPONENT | 15 | 4.62 | 2.15 |

Table 14: Specifications of a regression model at the preposition level, based on the **Belgian** data.

The measure VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION was predicted to be positively correlated with the probability of the prepositional variant. Figure 7 shows that this is not the case, neither in the Netherlands, nor in

Belgium. Instead, we find a non-significant negative effect in both countries. There are four possible reasons for this failed prediction, which we list from closest to furthest from our hypothesis. First, perhaps VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION is in fact positively correlated with the probability of the prepositional variant, but one or a few verbs are acting as outliers and jumble up the correlations in Figure 7. It was already suspected that *opbellen* 'phone' may act as such an outlier (see Section 5.2.12). Second, perhaps the verbs do exhibit lexical biases that are driven by some general semantic distinction, but the measure VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION has failed to capture the relevant semantic dimension. Third, perhaps the verb exhibit lexical biases that are, however, not driven by a general semantic distinction at the level of the preposition, but rather by distinctions at lower levels, or their lexical biases are even completely random. Fourth, perhaps the verbs exhibit no lexical biases at all.



a.    The Netherlandic model                    b.    The Belgian model
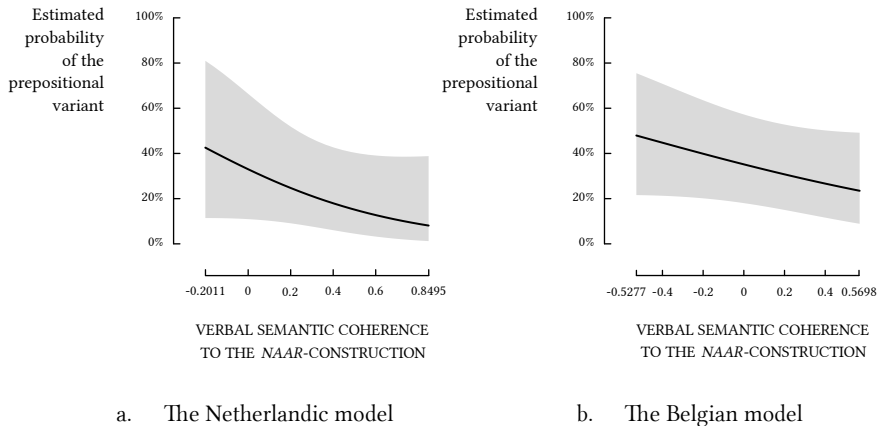
Figure 7: Effect plots of the Netherlandic and Belgian measures of VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION in respectively the Netherlandic model of Table and Belgian model of Table 14.

To find out which is the case, VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION is removed from the models, and VERB is implemented as a fixed effect. In this way, we can directly assess the lexical biases of the verbs in an effect plot, while controlling for differences in complexity and register. The specifications of the new regression models can be found in Table 24 and 25 in the Appendix. Multicollinearity was again not found to be a problem, with all VIFs well below 5 (Levshina 2015: 160). The Belgian model scores a C-index of 0.700, the Netherlandic model one of 0.744. Figure 8 shows the effect plot for VERB in the

Netherlandic and Belgian models. The verbs are ordered according to increasing probability of the prepositional variant in Belgium.
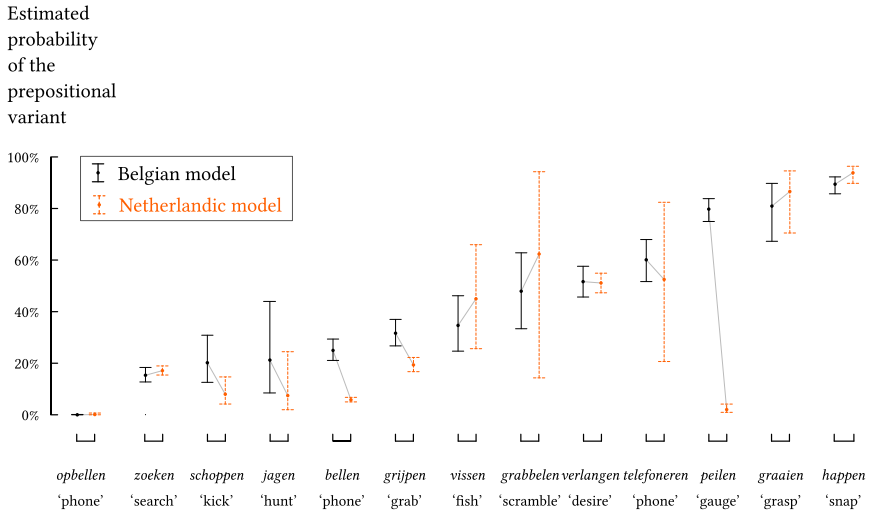


Figure 8: Effect plot of VERB in the Netherlandic model in Table 24 in the Appendix and the Belgian model in Table 25 in the Appendix.

The preferences for each variant do indeed strongly differ from verb to verb, but no general semantic distinction seems apparent from Figure 8, neither in Belgium, nor in the Netherlands. This means that either the distributions between the transitive and prepositional variant of a verb are determined by something else or they are completely random. Be that as it may, we can conclude that prediction Sem1.1 is not confirmed: it does not seem the case that verbs whose lexical meaning implies more directionality have an increased probability of the prepositional variant among their interchangeable instances.

We now evaluate the lectal hypotheses at the level of the preposition. Hypothesis Lec1 is clearly confirmed. Both Netherlandic models score an outspokenly higher C-index (0.745 and 0.744) than the Belgian models (both: 0.700).[63] Lec2 is also confirmed. The fixed effect VERB causes a greater increase in the predictive quality of the Netherlandic model (C-index + 0.109) than in the Belgian model (C-index + 0.079). The cause of this can readily be inferred from

---

[63] The C-index of the Belgian model with VERBAL COHERENCE TO THE NAAR-CONSTRUCTION as fixed effect and verb as random effect is actually slightly higher than its counterpart, but not when the C-indexes are rounded to 3 decimals.

Figure 8: the verbs generally seem to exhibit more outspoken lexical biases in the Netherlands than in Belgium. Especially for *peilen* 'gauge', we find a massive difference between the Netherlandic and Belgian regiolects. This does not come as a real surprise, since we had already found the prepositional variant of *peilen* 'gauge' to be too infrequent in the Netherlands to run collostructional analyses in Subsection 4.2.2.2.2.

## 6.2 Level of the verb

This section evaluates the lectal and semantic predictions at the level of the verb. We follow the same order as in Section 4.1: first the verb *verlangen* 'desire', then *peilen* 'gauge' and finally *zoeken* 'search'.

### 6.2.1 *Verlangen* 'desire'

For the analyses presented in this subsection, the data is restricted to the instances of the verb *verlangen* 'desire'. We compose a regression model with the fixed effects SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION and OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION (see Equation 2 in Section 4.2). Since these measures could only be calculated for full nominal theme roots, all instances with pronominal theme roots are removed from the dataset, as well as all instances with theme roots that did not appear with any of the 5000 most frequent context features in the corpus, since no vector could be calculated for them. Finally, the instances of the collexemes themselves were removed to avoid circularity.[64]

Because all instances of the same theme root had the same value for SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION and OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION, random intercepts were added for THEME ROOT. The variables THEME COMPLEXITY, VERB-THEME ORDER and an interaction between both are again added as fixed effects. In order to get both models to converge, all theme roots with only a single occurrence in the data had to be binned into a rest category for THEME ROOT (Wolk et al. 2013: 399) and the random effect for CORPUS COMPONENT had to be left out.[65]

---

[64] Diminutives have different ROOT-tags than their base forms in the Alpino-parses, but the diminutive forms of the collexemes were also removed.

[65] It was attempted to fit models with more coarse-grained classifications of the corpus components, but this was to no avail.

The measures SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION and OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION both have *huis* 'house' among the top 5 collexemes of the prepositional variant. Hence, we expected them to correlate with one another in the Netherlands and Belgium. This was indeed the case, with a Pearson correlation coefficient of 0.68 (p < 0.0001) for the Netherlandic measures and 0.85 for the Belgian measures (p < 0.0001).[66] Still, when both measures were put into their regression models, their respective VIFs were only 1.58 and 1.53 in the Netherlandic model and 3.65 and 3.59 in the Belgian model.[67] This is fairly high for the Belgian model, where the correlation between both measures is strongest, yet the values are still below most conventional thresholds (Levshina 2015: 160). As such, the measures were both retained in the Netherlandic and Belgian regression models. The other VIFs were all lower.

The specifications of the Netherlandic and Belgian models are listed respectively in Table 15 and Table 16. Both models exhibit excellent predictive quality, with C-indexes of 0.895 for the Netherlandic model and 0.840 for the Belgian model. Figure 9 shows the effect plots of SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION and OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION in the Netherlandic model, while Figure 10 does the same for the Belgian model.[68]

According to Prediction Sem2.1, we expected a positive correlation between SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION and the probability of the prepositional variant. We find this prediction to be clearly confirmed, both in the Netherlands and in Belgium. Meanwhile, Prediction Sem1.2 foresaw a positive correlation between OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION and the probability of the prepositional variant. This is clearly refuted. In the Netherlandic model, the measure does not appear to have any effect, and in the Belgian model, the effect is even negative. This negative effect is surprising, especially considering the strong positive correlation between the Belgian measures SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION and OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION. When SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION is removed from the model,

---

[66] In the calculation of these correlation strengths, each 'observation' corresponds to a single theme root. The correlations between the Netherlandic measures was calculated based on the Netherlandic data and the correlation between the Belgian measures on the Belgian data.

[67] As before, these VIFs were calculated in the models without the interaction between THEME COMPLEXITY and VERB-THEME ORDER.

[68] The Netherlandic model is based on more observations than the Belgian model, and as a result, THEME ROOT has more levels in the Netherlandic model than in the Belgian model. Since the number of levels scales with the number of observations, however, the explanatory power of THEME ROOT is not necessarily greater in the Netherlandic model than in the Belgian model.

OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION does have a clear positive effect.[69]

AIC: 894.7          Transitive observations: 448
C-index: 0.895      Prepositional observations (success level): 504

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | -0.76 | 0.30 | -2.58 | 0.0098 |
| COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION | | 3.32 | 0.33 | 10.17 | < 0.0001 |
| OBJECTAL COHERENCE TO THE *NAAR*-CONSTRUCTION | | -0.04 | 0.20 | -0.20 | 0.8413 |
| THEME COMPLEXITY | | -0.38 | 0.20 | -1.89 | 0.0587 |
| VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 1.61 | 0.32 | 4.98 | < 0.0001 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.06 | 0.24 | 0.26 | 0.7948 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| THEME ROOT | 154 | 1.58 | 1.26 |

Table 15: Specifications of a regression model fitted on the **Netherlandic** instances of the verb *verlangen* 'desire'.

AIC: 715.4          Transitive observations: 266
C-index: 0.840      Prepositional observations (success level): 411

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | 0.82 | 0.32 | 2.55 | 0.0108 |
| COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION | | 2.38 | 0.32 | 7.47 | < 0.0001 |
| OBJECTAL COHERENCE TO THE *NAAR*-CONSTRUCTION | | -0.85 | 0.24 | -3.57 | 0.0004 |
| THEME COMPLEXITY | | -0.83 | 0.26 | -3.20 | 0.0014 |
| VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.93 | 0.36 | 2.55 | 0.0109 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.44 | 0.31 | 1.42 | 0.1551 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| THEME ROOT | 105 | 1.43 | 1.20 |

Table 16: Specifications of a regression model fitted on the **Belgian** instances of the verb *verlangen* 'desire'.

---

[69] These results bear testimony to the importance of multifactorial testing.
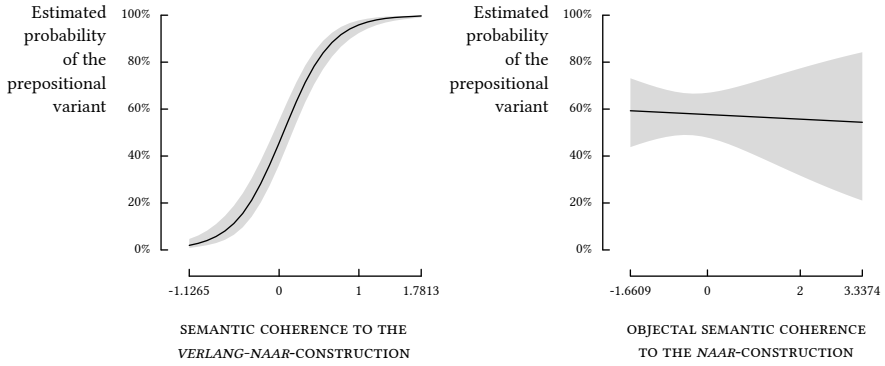
Figure 9: Effect plots of the Netherlandic model in Table 15.
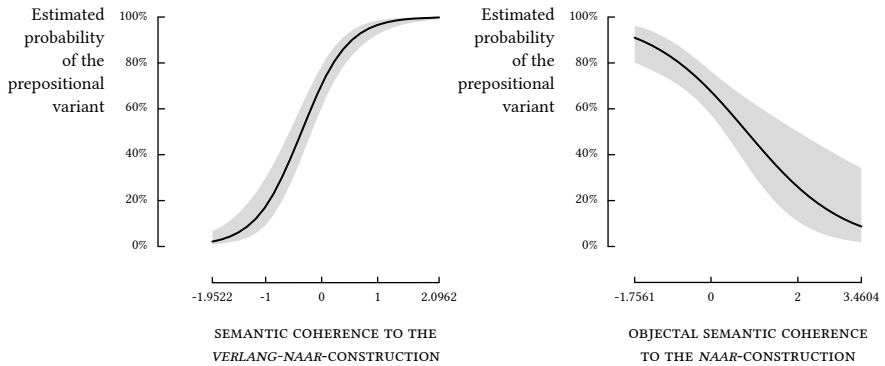


Figure 10: Effect plots of the Belgian model in Table 16.

The discrepancy between the Netherlandic and Belgian model regarding the influence of OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION seems to be due to the different collexemes used in the measures. When the collexemes *België* 'Belgium', *probleem* 'problem' and *werk* 'work', i.e. three Belgian collexemes that are not among the Netherlandic collexemes (see Tables 3-4 in Subsection 4.2.2.3.4), are removed from the calculation of the Belgian OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION, the measure no longer yields a significant effect. It seems that because of these three, the Belgian OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION is picking up on some signal leftover by SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION. Possibly,

people also tend to long for themes that are semantically similar to problems and work.[70] Be that as it may, we clearly have to conclude that Sem1.2 is not confirmed.

Turning to the lectal hypotheses, we again find Lec1 to be confirmed. The Netherlandic model has a higher predictive quality (C-index = 0.895) than the Belgian model (C-index = 0.840). As for Lec2, the variables that relate to lexical biases, viz. the fixed effects SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION and OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION and the random effect THEME LEMMA, cause an increase of 0.202 to the C-index of the Netherlandic model, versus an increase of 0.199 to the C-index of the Belgian model. The increase for the Netherlandic model is indeed greater, if only slightly. Lec2 is hence also confirmed.

In fact, the differences in the C-indexes and in the increase in C-index are larger when we leave OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION out of the models. There is good reason for this, because we do not really know what is causing its negative effect in the Belgian model. Concretely, the Netherlandic model without OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION still scores a C-index of 0.895, while the Belgian model only reaches 0.831. Regarding the increase in C-index, when we ignore OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION and compare models with SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION and THEME ROOT to models without these variables, then we still find an increase of 0.202 for the Netherlandic model, versus an increase of only 0.190 for the Belgian model.

## 6.2.2 *Peilen* 'gauge'

For *peilen* 'gauge', we had found that the prepositional variant was nearly non-existent in the Netherlands, while it was by far the most dominant variant in Belgium (see Figure 8). As a result, we cannot fit a regression model on the Netherlandic data.[71] Table 17 shows the specifications of the Belgian model. The same instances had to be excluded as for *verlangen* 'desire', mutatis mutandis. Instances of the overlapping collexemes *reactie* 'reaction' and *mening* 'opinion' were not removed, however, since these had no influence on SEMANTIC COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION (see Equation 3 in Subsection 4.2.2.2.2).

---

[70] Antonyms often occur in similar textual contexts, and distributional vector modelling therefore tends to rate them as semantically close (Turney and Pantel 2010: 149–150). Perhaps people often long for the antonyms of problems and work.

[71] A general rule of thumb is not to include any more parameters in the model than the number of instances of the least frequent response level divided by 20 (Speelman 2014: 530). Our Netherlandic dataset does not even contain sufficient instances of the prepositional variant for the inclusion of a single parameter.

In order to get the model to converge, all theme roots that occurred only once in the dataset again had to be binned into a rest category for the random effect THEME ROOT (Wolk et al. 2013: 399), but the random effect CORPUS COMPONENT could be retained in the model. All VIFs were well below 5, and the model has an excellent predictive quality with a C-index of 0.825. Figure 11 shows the effect plots of SEMANTIC COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION and OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION.

AIC: 510.9            Transitive observations: 116
C-index: 0.825        Prepositional observations (success level): 721

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | 0.01 | 0.48 | 0.03 | 0.9762 |
| SEMANTIC COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION | | 0.94 | 0.52 | 1.80 | 0.0724 |
| OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION | | 0.01 | 0.32 | 0.03 | 0.9786 |
| THEME COMPLEXITY | | -0.08 | 0.25 | -0.32 | 0.7526 |
| VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 2.52 | 0.60 | 4.21 | < 0.0001 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.32 | 0.37 | 0.86 | 0.3913 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| THEME ROOT | 124 | 0.87 | 0.93 |
| CORPUS COMPONENT | 11 | 0.05 | 0.23 |

Table 17: Specifications of a regression model fitted on the **Belgian** instances of the verb *peilen* 'gauge'.
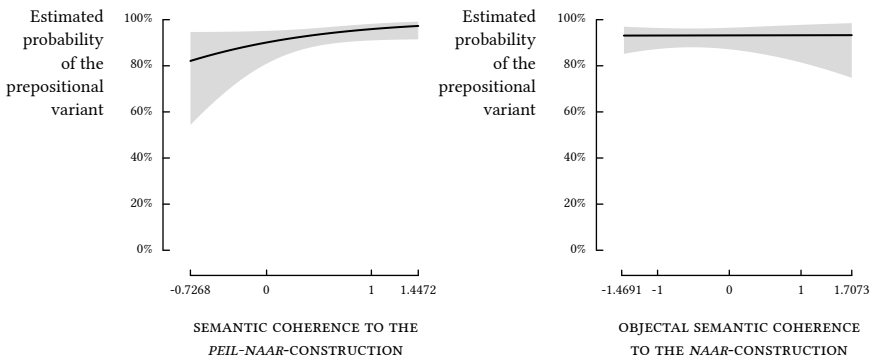


Figure 11: Effect plots of the **Belgian** model in Table 17.

Prediction Sem3.1 required a positive effect of SEMANTIC COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION on the probability of the prepositional variant. We do find such a positive effect, but it is non-significant, if only barely (p = 0.0724). This non-significance might be due to a ceiling effect, given the dominance of the prepositional variant in Belgium, and/or due to a simple lack of data for a comparatively complex model. Still, this means we cannot confirm prediction Sem3.1. Meanwhile, prediction Sem1.2 called for a positive effect of OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION on the use of the prepositional variant. This is clearly not confirmed.

### 6.2.3    *Zoeken* 'search'

We finally turn to the verb *zoeken* 'search'. The same instances that had to be excluded for *verlangen* 'desire' and *peilen* 'gauge', were also excluded for *zoeken* 'search'. Instances of the overlapping collexeme *oplossing* 'solution' were retained like they were for *peilen* 'gauge'.

   In order for the models to converge, their random structure again had to be simplified. All theme roots that only occurred twice or once in the datasets were binned into a rest category, and the corpus components were grouped into the following more general categories: *WR-E-nontechnical* (written to be read, electronic, non-technical: e-magazines, subtitles, teletext pages, web sites, blogs), *WR-P-nontechnical* (written to be read, printed, non-technical: books, brochures, printed newsletters, newspapers, periodicals and magazines, written assignments), *WR-P-technical* (written to be read, printed, technical: guides and manuals, legal texts, policy documents, proceedings, reports), and *WS* (written to be spoken: auto cues, texts for the visually impaired, see Table 1 in Section 3.1, Oostdijk et al. 2013b: 22).

   All VIFs were well below 5, and the predictive quality of both models is excellent, with C-indexes above 0.8 (Hosmer and Lemeshow 2000: 162). Table 18 and Table 19 show the specifications of respectively the Netherlandic and Belgian models, and Figure 12 and Figure 13 show the respective effect plots of SEMANTIC COHERENCE TO THE *ZOEK-NAAR*-CONSTRUCTION and OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION.

   Prediction Sem4.1 expected a positive correlation between SEMANTIC COHERENCE TO THE *ZOEK-NAAR*-CONSTRUCTION and the probability of the prepositional variant. We find this prediction to be clearly confirmed, both in the Netherlands and in Belgium. Meanwhile, we find a significant negative effect for OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION. It is hard to ascertain what is causing this, but prediction Sem1.2, which foresaw a positive effect, clearly has to be refuted. As for the lectal predictions, we again see a higher C-index for the Netherlandic model than for the Belgian model, confirming Lec1. This time,

however, we do find that the semantic and lexical variables cause a slightly higher increase in C-index for the Belgian model (C-index + 0.192) than for the Netherlandic model (C-index + 0.171). As a result, Lec2 cannot be confirmed here.

AIC: 11,292.2          Transitive observations: 10,875
C-index: 0.816         Prepositional observations (success level): 2,570

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | -2.16 | 0.15 | -14.13 | < 0.0001 |
| SEMANTIC COHERENCE TO THE ZOEK-NAAR-CONSTRUCTION | | 1.42 | 0.15 | 9.66 | < 0.0001 |
| OBJECTAL SEMANTIC COHERENCE TO THE NAAR-CONSTRUCTION | | -0.31 | 0.06 | -5.34 | < 0.0001 |
| THEME COMPLEXITY | | -0.37 | 0.06 | -6.32 | < 0.0001 |
| VERB-THEME ORDER | verb-theme (vs. theme-verb) | 0.51 | 0.08 | 6.15 | < 0.0001 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | verb-theme (vs. theme-verb) | 0.53 | 0.07 | 7.70 | < 0.0001 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| THEME ROOT | 734 | 1.41 | 1.19 |
| CORPUS CATEGORIES | 4 | 0.04 | 0.20 |

Table 18: Specifications of a regression model fitted on the **Netherlandic** instances of the verb *zoeken* 'search'.

AIC: 28,472.9          Transitive observations: 29,861
C-index: 0.808         Prepositional observations (success level): 6,161

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | -2.33 | 0.21 | -11.11 | < 0.0001 |
| SEMANTIC COHERENCE TO THE ZOEK-NAAR-CONSTRUCTION | | 1.49 | 0.14 | 10.78 | < 0.0001 |
| OBJECTAL SEMANTIC COHERENCE TO THE NAAR-CONSTRUCTION | | -0.23 | 0.05 | -4.82 | < 0.0001 |
| THEME COMPLEXITY | | -0.45 | 0.04 | -11.28 | < 0.0001 |
| VERB-THEME ORDER | verb-theme (vs. theme-verb) | 0.26 | 0.06 | 4.77 | < 0.0001 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | verb-theme (vs. theme-verb) | 0.57 | 0.05 | 12.36 | < 0.0001 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| THEME ROOT | 1,535 | 1.39 | 1.18 |
| CORPUS CATEGORIES | 4 | 0.14 | 0.37 |

Table 19: Specifications of a regression model fitted on the **Belgian** instances of the verb *zoeken* 'search'.
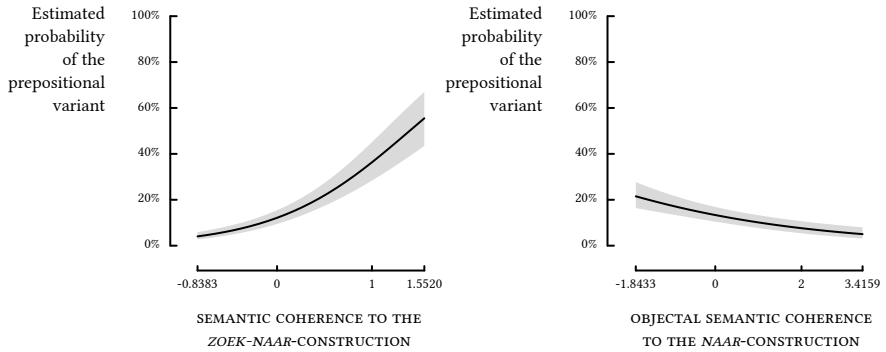
Figure 12: Effect plots of the **Netherlandic** model in Table 15.
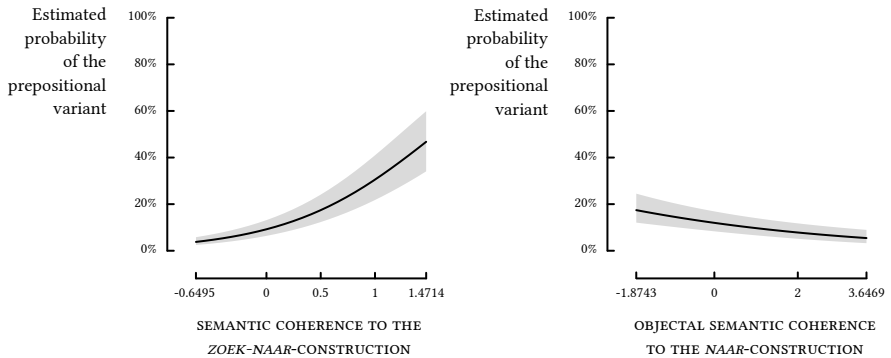


Figure 13: Effect plots of the **Belgian** model in Table 16.

## 6.3    Level of the object

The predictions Sem2.2, Sem3.2 and Sem4.2 envision differences among instances of a specific verb and a specific theme root. In other words, they pertain to the level of the object. The following subsections test these predictions, starting with the verb *verlangen* 'desire' and the theme roots *ding* 'thing' and *tijd* 'time', followed by the verb *peilen* 'gauge' and the theme roots *reactie* 'reaction' and *stemming* 'mood', and finally the verb *zoeken* 'search' and the theme roots *slachtoffer* 'victim' and *woord* 'word'.

This means the datasets are limited to the relevant instances, e.g. all instances that have *zoeken* 'search' in their verb slot and and *slachtoffer* 'victim' in their object slot. These instances were then annotated according to the criteria set in Section 4.4. During this annotation, the instances were blinded for the choice of

variant by removing the preposition *naar* 'to' from the instances of the prepositional variant, such that all instances appeared as though they were transitive. This was done to preclude any possibly subconscious confirmation bias. The label *unclear* was assigned when the appropriate label was not clear from the immediate context, viz. the sentence itself or the immediately preceding or following sentences.

### 6.3.1 *(Naar) ding en tijd verlangen 'desire thing and time'*

For *(naar) ding verlangen* 'desire thing', it was decided in Subsection 4.2.2.3 to attempt to annotate directly for the meanings 'demand' vs. 'long for'. The prediction was that instances where the meaning of *verlangen* 'desire' inclines more towards demand, as in (174a-b), would prefer the transitive variant compared to instances where the meaning inclines more towards 'long for', as in (174c-d).

Figure 14 contains a mosaic plot of the results. The width of the bar in these plots is proportional to the distribution of the instances over the labels, i.e. over the levels of the independent variable, while their height is proportional to the distribution of the instances over the variants, i.e. the levels of the dependent variable. The number in each 'cell' indicates the number of instances. The numbers in these cells are too low to seriously perform any kind of inferential statistics, but the results are in line with our hypothesis.

As for *(naar) tijd verlangen* 'desire time', prediction Sem2.2b stated that when time functioned as a mass noun, as in (175a-b), the transitive variant would be preferred, while when it functioned as a count noun, as in (175c-d), the prepositional variant would be used. The results are shown in the mosaic plot in Figure 15. Again, the data are extremely few in number, but they are in line with the hypothesis.

174. a. *Mannen verlangen eigenlijk maar drie dingen van een auto: een*
Men desire actually just three days of a car: a
*krachtige motor, (...)*
powerful engine
'Men really only desire three things from a car: a powerful engine, (...)
(WR-P-P-H-0000087207.p.12.s.1)

b. *Zo'n man verlangt naar kleine dingen: dat hij hartelijk wordt*
such_a man desires to small things that he cordially is
*ontvangen; (...)*
received
'Such a man desires little things: that he is cordially received (...)"
(WR-P-P-H-0000015490.p.7.s.4)

c.  *(...)  dat  wij turend naar de wereldbol maar  één  ding verlangden:*
           that we gazing to  the globe   just   one thing desired
*ooit  in  de lift   in  São Paulo te  staan – (...)*
ever in  the elevator in  São Paulo to  stand
'(...) that we, gazing at the globe, just desired one thing: to one day be standing in the elevator in Sao Paulo – (...)'
        (WR-P-P-G-0000190922.p.6.s.3)

d.  *Ongeacht   hoe sterk   we naar deze dingen verlangen, zowel  de*
    regardless_of how strongly we to  these things desire   both   the
*persoonlijke als de onpersoonlijke, (...)*
personal   as the impersonal
'Regardless of how strongly we desire these things, both the personal and the impersonal, (...)'
        (WR-P-P-B-0000000016.p.544.s.2)

175. a.  *Ik wil  niet sterven, maar ik mag niet nog meer tijd  verlangen*
       I  want not  die    but I  may not still more time desire
*dan ik al    heb geleefd.*
than I  already  have lived
'I do not want to die, but I may not desire even more time than I've already lived.'
        (WR-P-P-B-0000000183.p.393.s.2)

b.  *(...) als je   naar vrije tijd  verlangt.*
       if  you to   free time desire
'(...) if you desire some time off.'
        (WR-P-P-G-0000602126.p.7.s.3)

c.  *Ik verlangde weer  naar een  rustige  tijd  met onze wandelingen*
    I  longed    again to  a    peaceful time with our  walks
*samen.*
together
'I longed, once more, for a peaceful time with our walks together.'
        (WR-P-P-B-0000000021.p.1303.s.8)

d.  *Als (...), verlangt ze   soms    naar de tijd  dat  mensen*
    when   desires she sometimes to  the time that people
*zich     zonder 'al dat lawaai' konden amuseren.*
themselves without all that noise   could   amuse
'When (...), she sometimes longs for the time when people could amuse themselves without "all that noise"'
        (WR-P-P-H-0000098086.p.5.s.8)
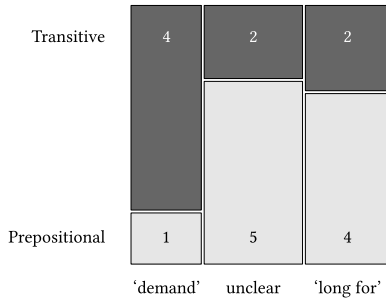
Figure 14: Mosaic plot of the manual annotation on *(naar) ding verlangen* 'desire thing'
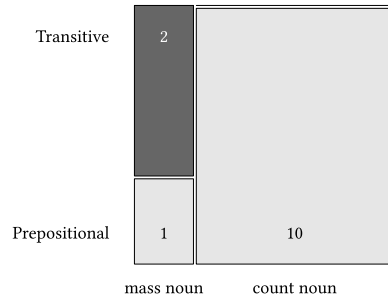


Figure 15: Mosaic plot of the manual annotation on *(naar) tijd verlangen* 'desire time'

## 6.3.2 *(Naar) reactie en stemming peilen* 'gauge mood and reaction'

To test predictions Sem3.2a-b, all instances of the verb *peilen* 'gauge' and the theme roots *reactie* 'reaction' and *stemming* 'mood' were manually marked as to whether the agent was directly judging the reactions or the mood him or herself, or was merely asking about the reactions or the mood. For direct judgments, as in (176a-b) for *reactie* 'reaction' and (177a-b) for *stemming* 'mood', the use of the transitive variant was predicted, whereas for gauging by asking questions, as in (176c-d) and (177c-d), the prepositional variant would be preferred.

Figure 16 and Figure 17 show the results. For *(naar) reactie peilen* 'gauge reaction', we find a significant difference between those instances marked as 'judging' and those marked as 'asking' (Fisher's exact test: p = 0.0002). For *(naar) stemming peilen* 'gauge mood', we only dispose of 6 instances that could be clearly marked as either category, but the results are at least in line with the hypothesis.


176. a. *(...) persvertegenwoordigers kijken ons  aan  om onze reacties    te*
          press_representatives      look    us    part to  our   reactions  to
      *peilen.*
      gauge
      '(...) press representatives look at us to gauge our reactions.'
              (WR-P-P-G-0000352753.p.9.s.4)

b. *Nu ga ik in de zaal en peil ik naar de reacties bij de*
now go I in the hall and gauge I to the reactions with the
*toeschouwers.*
spectators
'Now I go into the hall and I gauge the reactions among the
spectators.'
(WR-P-P-G-0000540751.p.6.s.2)

c. *Het Eén-magazine Koppen zendt het gesprek uit en*
the Eén-program Koppen broadcasts the conversation PART and
*ging de reacties peilen van Vlaamse betrokkenen. "Een*
went the reactions gauge of Flemish those_involved an
*adoptiekind ter plekke uitkiezen, heeft iets van een*
adoption_child at_the place pick_out has something of a
*slavenmarkt", vindt Jean Bosco Safari, (...)*
slave_market finds Jean Bosco Safari
'The Eén-program Koppen broadcasts the conversation and went out
to gauge the reactions of the involved Flemings. "Chosing an adoption
child on the spot resembles something of a slave market", feels Jean
Basco Safari (...)'
(WR-P-P-G-0000262161.p.2.s.3)

d. *Na de eerste aflevering van 'Aspe' op VTM peilde de*
after the first episode of Aspe on VTM gauged the
*krant naar reacties in Brugge. De meningen zijn verdeeld.*
newspaper to reactions in Bruges the opinions are divided
'After the first episode of Aspe on VTM, the newspaper gauged the
reactions in Bruges. The opinions are divided.'
(WR-P-E-C-0000014816.p.1.s.6)

177. a. *Volgens Salam Pax, (...) zijn blogs de efficiëntste manier*
according_to Salam Pax are blogs the most_efficient way
*om de stemming in Irak te peilen.*
to the mood in Iraq to gauge
'Accoding to Salam Pax, (...) blogs are the most efficient way to gauge
the mood in Iraq.'
(WR-P-P-G-0000227458.p.3.s.1)

b. *Via hen kan hij beter de stemming op de werkvloer peilen.*
via them can he better the mood on the shop_floor gauge
'Via them, he is better able to gauge the mood on the shop_floor.'
(WS-U-E-A-0000121674.p.1.s.12)

c.   *Hij trekt   van Lahore   naar de Swatvallei,   praat   met   politici, (...)*
he  travels from Lahore   to     the Swat_valley talks   with  politicians
*en    peilt    de stemming bij   de man in de straat.*
and   gauges the mood        with the man in  the street
'He travels from Lahore to the Swat valley, talks to politicians, (...) and
gauges the mood with the man in the street.'
         (WR-P-P-H-0000121429.p.1.s.2)

d.   *Net   voor   het dieptepunt  peilden  we al         tussentijds       naar*
just  before the low_point   gauged  we already  between_times  to
*de    stemming bij   de 4 stockpickers (...), maar ze   hielden  het*
the   mood        with the 4 stockpickers        but  they held      the
*hoofd   terecht   koel. De overwegende reactie   was: "een kans   om*
head    rightly    cool  the predominant  reaction was   a      chance to
*(bij)    te   kopen".*
(with)   to   buy
'Just before the low point, we already gauged the reaction of the
stockpickers between times (...), but they rightly kept their heads cool.
The predominant reaction was: "a chance to buy (extra)".'
         (WR-P-P-H-0000110893.p.4.s.1)
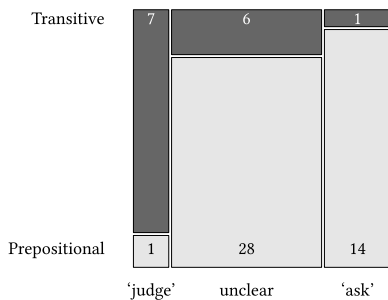


Figure 16: Mosaic plot of the manual
annotation on *(naar) reactie peilen* 'gauge
reaction'.
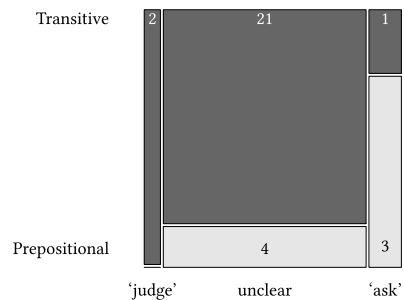


Figure 17: Mosaic plot of the manual
annotation on *(naar) stemming peilen*
'gauge mood'.

### 6.3.3 *(Naar) slachtoffer en woord zoeken* 'search victim and word'

The shortage of data is less strangling for *(naar) slachtoffer zoeken* 'search victim' and *(naar) woord zoeken* 'search word'. For *(naar) slachtoffer zoeken* 'search victim', Sem4.2a predicted a preference for the transitive variant when the agent is an aggressor, as in (178a-b), and a preference for the prepositional variant when the agent is a helper, as in (178c-d). All instances *(naar) slachtoffer zoeken* 'search victim' were marked for this distinction, and those where it was unclear were removed from the data. This left us with 68 instances of the least frequent variant, which is sufficient for a regression model with 3 parameters, following the rule of thumb mentioned in Footnote 71.

A regression model was composed with the fixed effects AGENT-TYPE (aggressor, helper), COUNTRY (Belgium, the Netherlands) and THEME COMPLEXITY. All VIFs were well below 5, and the model has excellent predictive quality with a C-index of 0.846. The specifications of this model can be found in Table 26 in the Appendix, and the effect plot of AGENT-TYPE is shown in Figure 18. We find that the prediction is clearly confirmed. Upon closer inspection, the effect might actually be even stronger than it appears in Figure 18. 14 of the instances stem from headlines, which were decided to be retained in the dataset in Subsection 5.2.1.2. 12 of these headline instances exhibit the transitive variant, while only 2 display the prepositional variant. Removing the headlines from the data raises the estimate for AGENT-TYPE to 4.07 ($p < 0.0001$).

When we look at the instances that go against the aggressor vs. helper distinction, i.e. aggressor instances that exhibit the prepositional variant and helper instances that exhibit the transitive variant, then a disproportionate number of them stems from Belgium. This could indicate that the distinction is more strict in the Netherlands than in Belgium, which would be in line with the second lectal hypothesis. This difference is not significant, however.[72]

178. a.    *De potentiële zakkenrollers zoeken waarschijnlijk een nieuw*
           the potential pickpockets   search probably         a     new
           *slachtoffer.*
           victim
           'The potential pickpockets are probably already looking for a new victim.'
                    (WR-P-P-G-0000140427.p.33.s.2)

---

[72] Fisher's exact test including the headlines: $p = 0.1996$, Fisher's exact test excluding the headlines: $p = 0.3628$. These test where executed on the correlation between the variables INLINE-WITH-DISTINCTION (*inline, not-inline*) and COUNTRY.

b.  *De politie vreest dat de daders      buitenlanders zijn die in*
    the police fears  that the perpetrators foreigners    are  that in
    *Brussel op rooftocht waren. «Ze zoeken op het spitsuur in het*
    Brussels on raid       were  they search  on the rush_hour in the
    *station naar een  slachtoffer»*
    station to    a    victim.'
    'The police fears that the perpetrators were foreigners that were doing
    a raid in Brussels. "They look for victims during rush hour in the
    station.'
             (WR-P-P-G-0000357007.p.3.s.2)

c.  *Eén van de stelregels als   je   een slachtoffer van verdrinking*
    one of  the maxims  when you a   victim     of   drowning
    *zoekt... is dat je   de hoop nooit mag opgeven.*
    search is  that you the hope never may give_up
    'One of the maxims when you are looking for a victim of a drowning...
    is that you should never give up hope.'
             (WR-P-P-G-0000357007.p.3.s.2)

d.  *In Albanië zijn honderden hulpverleners aan het werk op de*
    In Albania are hundreds  relief_workers on  the work on the
    *plek  van de zware ontploffingen. Ze  zoeken naar slachtoffers*
    place of the heavy explosions     they look   for  victims
    *en   naar niet-ontplofte munitie.*
    and to   not-exploded munitions
    'In Albania, hundreds of relief workers are working on the place of the
    heavy explotions. They are looking for victims and for munitions that
    haven't exploded yet.'
             (WS-U-E-A-0000291696.p.1.s.2)
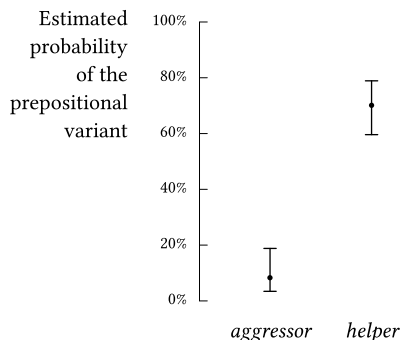


Figure 18: Effect plot of the manual annotation on *(naar) slachtoffer zoeken* 'search
victim', corresponding to the regression model in Table 26 in the Appendix.

We now turn to *(naar) woord zoeken* 'search word'. Here, Sem4.2b predicted a higher probability of the prepositional variant when the word at issue was specific, as in (179a-b) compared to when it was non-specific, as in (179c-d). Again, instances where this was unclear, were excluded from the dataset, which left us with 83 instances of the least frequent variant.

We then composed a regression model with as fixed effect WORD SPECIFICITY (*unspecific, specific*), COUNTRY, THEME COMPLEXITY and VERB-THEME ORDER. All VIFs were well below 5, and the C-index indicates acceptable discrimination, with C = 0.732. Table 27 in the Appendix shows the specifications of this model, and the effect plot of WORD SPECIFICITY can be found in Figure 19. We find exactly the opposite effect to what was predicted. Prediction Sem4.2b is clearly not confirmed. When we look at the instances that go against the distinction in Figure 19, we again find that they are disproportionally Belgian. This difference is again not significant, however (Fisher's exact test: p = 0.1344).

179. a.  *Het is zo... (zoekt   haar woorden) instant.*
        it    is    so    searches her   words      instant
        'It is so... (searches for words) instant.'
                (WR-P-P-G-0000178976.p.29.s.11)

    b.  *Dirk Elsen zoekt    naar de juiste  woorden. Zoiets     als*
        Dirk Elsen searches to   the right   words     something like
        *'geitenwollen sokken' wil    hij beslist   niet in de mond nemen.*
        goat's_whool socks     wants he definitely not  in the mouth take
        'Dirk Elsen searches for the right words. He definitely does not want to say domething like "open sandals and woolly socks type".'
                (WR-P-P-G-0000087838.p.40.s.1)

    c.  *Zoek in het  bronbestand het   woord  Amersfoort.*
        Search in the   source_file   the   word   Amersfoort
        'Look for the word Amersfoort in the source file.'
                (WS-U-T-B-0000000891.p.8.s.1)

    d.  *Er    werd gesleuteld tot   er    een show ontstond met een*
        there was    fiddled    until there a    show arose    with a
        *flinterdun    verhaal waarin je    tevergeefs naar de woorden "fuck"*
        paper-thin   story    wherein you vainly   to   the words   fuck
        *en   "cut" zal  zoeken.*
        and   cut   will   search
        'There was fiddling until a show emerged with a paper-thin story in which you will vainly look for the words "fuck" and "cut".'
                (WR-P-P-G-0000586752.p.3.s.1)
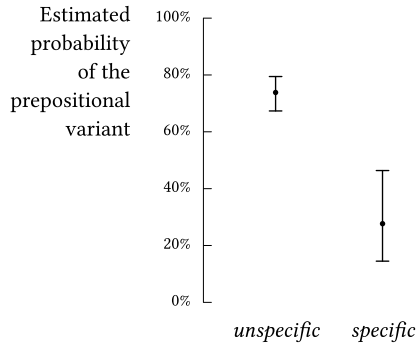
Figure 19: Effect plot of the manual annotation on *(naar) woord zoeken* 'search word', corresponding to the regression model in Table 27 in the Appendix.

## 6.4    Overview and interpretation

We find that the lectal hypotheses, repeated below, have generally been confirmed. All Netherlandic models scored higher C-indexes than their Belgian counterparts. This indicates that the transitive-prepositional alternation in Netherlandic Dutch is generally easier to model than in Belgian Dutch, or, more generally, that Belgian Dutch seems to behave more heterogeneously with respect to language variation than Netherlandic Dutch. As for the second lectal hypothesis, we find it to be confirmed for all models except for the ones fitted on the data of *zoeken* 'search'. Variation in the Netherlands indeed seems to be more strictly determined by lexical preferences and semantic distinctions.

The lectal hypotheses have been confirmed both at the level of the preposition and the verb. In other words, variation in argument structure appears to be more strongly determined by the choice of verb in the Netherlands, with each verb showing a more outspoken preference for either the transitive or the *naar*-construction. Meanwhile, when a Netherlandic verb does exhibit both variants in a balanced distribution – which is really only the case for *verlangen* 'desire' (cf. Figure 8)[73] – this variation is more strictly regulated by a semantic distinction. This confirms earliers findings by among others Grondelaers, Speelman and Geeraerts (2008).

Generally, the differences between the Belgian and Netherlandic regiolects tend to be larger in informal registers than in formal ones (Grondelaers, Speelman and

---

[73] The verbs *vissen* 'fish', *grabbelen* 'scramble' and *telefoneren* 'phone' also appear to exhibit fairly balanced distributions in the Netherlandic model in Figure 8, but these verbs display wide confidence intervals.

Geeraerts 2002; Daems, Heylen and Geeraerts 2015; Geeraerts 2017). Since the Sonar-corpus contains predominantly formal, written language, we could reasonably expect even larger differences when more informal material and possibly spoken language would be taken under scrutiny.

**Lec1**: The predictive quality of the models fitted on the Netherlandic data will generally be higher than those fitted on the Belgian data. **Confirmed.**

**Lec2**: The predictors relating to lexical biases will cause a greater increase in predictive quality in the Netherlandic models than in the Belgian models. **Generally confirmed.**

The semantic hypotheses are repeated below. Perhaps the most striking result is the complete failure of the predictions based on the hypothesis at the level of the preposition, i.e. Sem1. The reason of this failure becomes clear when surveying the results of Sem2, Sem3 and Sem4, which generally have been successful. When each verb employs the alternation to express its own meaning distinction, this inevitably goes at the cost of a more general, shared semantic contrast.

For example, the transitive-prepositional alternation appears to express the distinction between a desire construed as a demand and one construed as a longing for the verb *verlangen* 'desire'. If language users mean to construe a desire as a demand in approximately 50% of the instances, and as a longing in the other 50%, this results in a nicely balanced distribution between the variants for the verb *verlangen* 'desire'. Meanwhile, for the verb *zoeken* 'search', another distinction, e.g. between 'seek to make/acquire' and 'look for', would result in another distribution, viz. approximately 80%-20%. Crucially, the distributions of the variants among the instances of *verlangen* 'desire' and *zoeken* 'search' would be independent of one another. In that case, prediction Sem1.1, which required that the distribution between the variants would be more biased towards the transitive variant for *verlangen* 'desire' than for other verbs such as *zoeken* 'search', would be bound to fail.

**Sem1**: The *naar*-construction expresses directionality, the transitive construction does not.

　　　1.　The measure VERBAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION will be positively correlated with the probability of the prepositional variant in the Netherlands and Belgium. **Not confirmed.**

2. The measure OBJECTAL SEMANTIC COHERENCE TO THE *NAAR*-CONSTRUCTION will be positively correlated with the probability of the prepositional variant in the Netherlands and Belgium, among instances of a single verb. **Not confirmed.**

**Sem2**: The meaning of 'desire' has specialized to 'demand' for the transitive *verlang*-construction, and to 'long for' for the *verlang-naar*-construction.

1. The measure SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION will be positively correlated with the probability of the prepositional variant in the Netherlands and Belgium, among the instances of the verb *verlangen* 'desire'. **Confirmed.**

2a. In the Netherlands and Belgium, among instances of the verb *verlangen* 'desire' and the theme root *ding* 'thing', those instances that involve longing for a thing will prefer the prepositional variant, whereas those demanding a thing will prefer the transitive variant. **Confirmed, but we have to be careful due to data scarcity.**

2b. In the Netherlands and Belgium, among instances of the verb *verlangen* 'desire' and the theme root *tijd* 'time', those instances where *tijd* 'time' acts as a count noun will prefer the prepositional variant, whereas those where *tijd* 'time' is a mass noun will prefer the transitive variant. **Confirmed, but we have to be careful due to data scarcity.**

**Sem3**: The meaning of 'gauge' has specialized to 'directly judging' for the Belgian *peil*-construction, and to 'gauging by asking' for the Belgian *peil-naar*-construction.

1. The measure SEMANTIC COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION will be positively correlated with the probability of the prepositional variant in Belgium, among the instances of the verb *peil* 'gauge'. **Not confirmed, but might be due to a ceiling effect and data scarcity.**

2a. Among instances of the verb *peil* 'gauge' and the theme root *stemming* 'mood', those that involve asking about the mood will prefer the prepositional variant, whereas those where the mood is directly judged will prefer the transitive variant. **Confirmed, but we have to be careful due to data scarcity.**

2b. Among instances of the verb *peil* 'gauge' and the theme root *reactie* 'reaction', those that involve asking about a reaction will prefer the prepositional variant, whereas those where a reaction is directly judged will prefer the transitive variant. **Confirmed, but we have to be careful due to data scarcity.**

**Sem4**: The meaning of the transitive *zoek*-construction inclines more towards 'seeking to make or to acquire something', whereas the meaning of the *zoek-naar*-construction inclines more towards an act of literally looking for something.

1. The measure SEMANTIC COHERENCE TO THE *ZOEK-NAAR*-CONSTRUCTION will be positively correlated with the probability of the prepositional variant in the Netherlands and Belgium, among the instances of the verb *zoeken* 'search'. **Confirmed.**

2a. In the Netherlands and Belgium, among instances of the verb *zoeken* 'search' and the theme root *slachtoffer* 'victim', those instances where the agent is a helper will prefer the prepositional variant, whereas those where the agent is an aggressor will prefer the transitive variant. **Confirmed.**

2b. In the Netherlands and Belgium, among instances of the verb *zoeken* 'search' and the theme root *woord* 'word', those instances where the agent is searching for specific words will prefer the prepositional variant, whereas those where agent is searching for non-specific words, e.g. to try to express a proposition, will prefer the transitive variant. **Not confirmed.**

Usage-based theory expects language variation to reflect a functional need (Diessel 2015), but this functional need may manifest itself at any level of abstraction. Concretely, a prediction such as Sem1.2 only has a chance of success if the alternating verbs share a common interest in expressing one specific type of semantic contrast. However, a distinction in terms of e.g. directionality may be relevant for one verb, but less so or not at all for another. Similarly, it may be really useful to be able to distinguish between 'judging' and 'asking' for *peilen* 'gauge', but not at all for *verlangen* 'desire'. If there is no such functional need shared between the verbs, it is no wonder that the various verbs do not converge towards a shared goal, such as expressing a common meaning contrast.

When would it be useful for various verbs to express a single, common semantic contrast? One possible answer is that is more often the case for verbs that are semantically alike. For example, psych verbs such as *storen* 'disturb', *verbazen* 'amaze' and *ergeren* 'annoy' all express the mental state of an experiencer caused

by some stimulus. As a result, it could be useful for all of these psych verbs to be able to construe the expressed action as either involving a very agentive stimulus and a rather passive experiencer, or vice versa (Pijpops and Speelman 2017).

Still, it is difficult or perhaps even impossible to know a priori how semantically close verbs have to be in order to jointly express the same semantic contrast. For instance, the various verbs appearing in the English conative alternation may all seem semantically rather close, but still express different meaning distinctions varying from one verb class to another (Perek 2015: 105−144). Conversely, the English verbs *send* and *pour* as are semantically quite far apart, but still can both be argued to express the same contrast in the dative alternation, viz. a distinction between transfer of possession and material transfer, as in (180)-(181) (Langacker 1990: 13−14; Goldberg 1995: 141−151; Colleman and De Clerck 2009: 9−12).[74] As such, it was at least plausible to hypothesize that the transitive-prepositional alternation expressed the same semantic contrast even among semantically somewhat distant verbs such as *zoeken* 'search' and *peilen* 'gauge'.

180. a.   *Bill sent Joyce a walrus.*
     b.   *Bill sent a walrus to Joyce.*
                    (taken from Langacker 1990: 13, cited in Colleman and De
                    Clerck 2009: 10)

181. a.   *Jack poured Jane a martini.*
     b.   *Jack poured a martini to Jane.*
                    (adapted from Goldberg 1995: 147)

Of course, we have only looked in depth at three verbs so far, so we should not jump to conclusions. Four possible results remain for the other verbs. First, the verbs *verlangen* 'desire', *peilen* 'gauge' and *zoeken* 'search' may simply be the odd verbs out, and all of the other verbs might still share a single semantic contrast, either in terms of directionality or in other terms. After all, *verlangen* 'desire', *peilen* 'gauge' and *zoeken* 'search' were selected precisely because they seemed to be semantically isolated (cf. Subsection 4.2.2.2). If so, then we should expect the analyses in Chapter 7 to point to such a systematic difference.

Second, perhaps the subgroups of semantically similar verbs act as verb classes, whereby each class uses the alternation to express its own semantic contrast, tailored to the specific needs of that class. In Table 2 of Section 3.2, we distinguished

---

[74] For the dative alternation too, however, the exact semantic contrast that is expressed sometimes does appear to differ from one verb to another or from one verb class to another (Goldberg 1995: 141−179; Gries and Stefanowitsch 2004: 104−107; Colleman and De Clerck 2009).

between (i) the telephonic verbs, with the (near-)synonyms *telefoneren*, *opbellen* and *bellen* (all: 'phone'); (ii) the motoric verbs, with the near-synonyms *grijpen* 'grab', *grabbelen* 'scramble' and *graaien* 'grasp', as well as *happen* 'snap' and *schoppen* 'kick', all of which express a movement of a body part, and (iii) the venatic verbs *jagen* 'hunt' and *vissen* 'fish'. If each of these groups of verbs band together, then the analyses in Chapter 7 should yield consistent semantic contrasts among members of the same verb group. In that chapter, we will use a data-driven technique to track down semantic contrasts among individual verbs and even verb-theme combinations

Third, perhaps every semantic distinction should be sought at even lower levels of abstraction, and each verb or even each verb-theme combination uses the alternation to express their own idiosyncratic contrasts. Fourth, semantic factors might not be at play at all for the other verbs.

Another striking result is the failure of prediction Sem4.2b, given the success of Sem4.1 and Sem4.2a. Again, there are multiple possible reasons. First, perhaps another distinction is at play for the verb *zoeken* 'search', different from 'seek to make/acquire' vs. 'look for', that can account for both the results of Sem4.1, Sem4.2a and Sem4.2b. We currently do not see how to conceive such a distinction, however. Second, the object *woord* 'word' may be the odd one out. Perhaps *naar woord zoeken* developed as a construction in its own right, meaning 'attempt to express something', while most other objects of *zoeken* 'search' do exhibit a distinction in terms of 'seek to make/acquire' vs. 'look for'.

Third, perhaps groups of semantically similar objects act as object classes that each use the alternation with *zoeken* 'search' to express idiosyncratic semantic contrasts, specific to each object class. For instance, a group around *slachtoffer* 'victim' may use it to express a distinction between 'seek to make/acquire' vs. 'look for', while a group around *woord* 'word' may use it for something else. Fourth, perhaps each individual object actually uses the alternation to express their own individual contrasts, and the confirmation of Sem4.1 was merely due to chance. To find out which is the case, we need to investigate the alternation separately for many different each object fillers, i.e. theme roots, of the verb *zoeken* 'search'. For this, we need a hypothesis-generating technique that functions at concrete levels, is scalable and can work with limited amounts of data. Such a technique is proposed in Chapter 7.

To end Chapter 6, should we conclude that our results refute the lexical origin mechanism, given the failure of hypothesis Sem1? Not at all. The predictions of Sem2, Sem3 and Sem4 were just as well based on the lexical origin mechanism, but simply applied it to a different level of abstraction, viz. the level of the verb. These predictions were generally confirmed. If the lexical origin mechanism is already operative at the level of the verb, it effectively blocks off its own application to the level of the preposition. Put concretely, if each verb already developed a meaning distinction tailored to its own specific functional needs, then these various verb-

specific distinctions obstruct the development of a more general semantic contrast, as argued above for *verlangen* 'desire' and *zoeken* 'search'. These results are fully in line with the results of Perek (2015: 111) for the English conative alternation, who already argued that the lexical origin mechanism should not necessarily pertain to the highest level of abstraction.

# 7  Hypothesis-generation at low levels of abstraction

The previous chapter concluded by noting the need to investigate the alternation at low levels of abstraction. Once we put this attention for more concrete levels into practice, however, we run into a number of problems. First, we need hypotheses. If every theme class or every individual theme indeed employs the alternation to express a meaning contrast that is tailored to its own idiosyncratic needs, then it is very hard to devise beforehand what that contrast might be. Second, as we investigate ever more concrete constructions, the number of distinct constructions increases exponentially. There are simply much more concrete constructions than abstract ones, and hence it takes a lot of time to look at all of them individually. Third, even in spite of using one of the largest available corpora of Dutch, we will suffer from data scarcity as we restrict our datasets to specific verbs and specific themes.

We are thus in need of a method (i) that is data-driven, i.e. focuses on generating hypotheses, and casts its nets wide; (ii) that is easily scalable, such that many concrete constructions can be investigated with relative ease; and (iii) that can work with limited amounts of data. In what follows, we propose to employ Memory-based Learning (MBL) as such a technique.

This chapter is structured as follows. Section 7.1 explains why and how we intend to use Memory-based Learning as a hypothesis-generating technique, and then tests this technique on the case studies where we already know what results to expect, viz. the alternations at the verb-level of *verlangen* 'desire' and *peilen* 'gauge', and the verb-theme combinations *slachtoffer zoeken* 'search victim' and *woord zoeken* 'search word'.[75] It will be shown that the technique does point

---

[75] We will not use the alternation at the verb-level for *zoeken* 'search' to test the procedure, because we want to use much of its data in the hypothesis-generating and, crucially, the hypotheses-testing analyses at the level of the object (see below). Furthermore, the case

towards the relevant distinctions. Section 7.2 then applies the technique to the other verbs and verb-theme combinations that are instantiated in a minimum amount of corpus instances, and discusses what hypotheses could be formulated based on the MBL-analyses. Finally, Section 7.3 tests these hypotheses by manually annotating data that were kept out of the MBL-analyses, and interprets the results of this manual annotation. Finally, Section 7.4 concludes this chapter.

## 7.1 Using Memory-based Learning for hypothesis-generation

### 7.1.1 Introducing Memory-based Learning

Memory-based Learning is a k-nearest neighbor classifier that can predict the choice between two or more linguistic variants for a new data point, based on a number of training data. It does so by calculating the proximity of the new data point to all instances of its training data and then selects the k training observations that most closely resemble the new data point. Based on the choice of variant in these k observations, it finally predicts the variant used in the new data point. It is thus a form of 'lazy learning': it does not build a model on the training data, such as a regression formula or a classification tree. Of course, the researcher does need to specify a number of features on the basis of which the proximity between the new data point and the known observations is calculated.

On a conceptual level, Memory-based Learning can be seen as the implementation of exemplar-based models of language processing.[76] This is not directly at issue here, however: we mean to employ it as a machine learning technique that is suitable for our current purposes. More information on Memory-based Learning can be found in Daelemans and van den Bosch (2005), and more practical information on the Memory-based Learning implementation that is used in the present investigation, viz. the Timbl-classifier, is available in Daelemans et al. (2010). For other examples of the application of Memory-based Learning in fundamental linguistic research, see Scha, Bod and Sima'an (1999), Keuleers and

---

studies of *ding verlangen* 'desire thing', *tijd verlangen* 'desire time', *reactie peilen* 'gauge reaction' and *stemming peilen* 'gauge mood' are not used because these were too infrequent to reliably confirm a relevant distinction (cf. Section 6.3).

[76] As such, Memory-based Learning can be argued to represent cognitively real processes quite accurately, or at least more so than e.g. regression modelling (van den Bosch and Daelemans 2013).

Daelemans (2007), Theijssen (2012), van den Bosch and Daelemans (2013) and van den Bosch, Grondelaers and Speelman (forthc.).

The following properties of Memory-based Learning (MBL) make it useful for hypothesis-generation at low levels of abstraction. First, it allows for various features to be tested simultaneously and these features may be very diverse in nature. Second, the features may have a lot of levels. This means we do not need to a priori decide on any high-level independent variables. Third, the features may correlate with one another, without such multicollinearity causing problems. Fourth, because Memory-based Learning does not build any sort of model on the training data, it does not make any assumptions regarding the structure or distribution of the training data. All of this means that (i) we can cast our nets wide when searching for hypotheses; (ii) that the analyses can easily be automated, such that we can investigate a large number of concrete alternations; and (iii) that Memory-based Learning can work with a limited amount of data for each alternation.[77]

## 7.1.2    Defining the procedure

The data-driven procedure that we propose consists of the following three steps. The first step involves choosing the features that the MBL-classifier will use to predict the choice of variant. In the second step, we run the MBL-analyses on each of the alternations. When choosing which alternations to examine further, the C-index could be useful, as this is an indication of the predictive quality of the analyses that is comparable across different baselines (cf. Chapter 6). If it is low, it is likely that the MBL-analysis did not manage to pick up on anything that could be useful for predicting the choice of variant. In the third step, we turn to the features of the retained analyses, and check their gain ratios. These gain ratios signal how useful each feature was for predicting the variant: features with a high gain ratio are probably relevant for the alternation at issue (see Quinlan 1993 on the calculation of gain ratios). We then 'look under the hood' of these features, and try to determine what makes them so useful. Finally, we use this information to conceive a hypothesis about what might drive the alternation.

We run the MBL-analyses on two levels of abstraction, viz. the level of the verb and the level of the object. Two types of MBL-analysis are executed on each alternation, viz. window-based and parse-based analyses (cf. De Troij et al. forthc.). On the one hand, the window-based analyses simply take as features the five words left and right from start of the theme argument, i.e. the place where the preposition

---

[77] Granted, its results may not always remain very stable when they are based on only few data (cf. De Troij et al. forthc.).

would appear, if the prepositional variant is used.[78] Sentence boundaries were not crossed: if the sentence contained less than 5 words left or right from the start of the theme, the corresponding features would have level *x*.

On the other hand, the parse-based analyses use information that can be readily extracted from the Alpino-parses of the sentence. Their features are listed below. In addition, all analyses also take the lectal features COUNTRY, with levels *Belgium* and *the Netherlands*, and COMPONENT, with each corpus component as a separate level. The advantage of the window-based analyses is that they cast their the nets wider, in that they have the potential of detecting possibly distinctions that are not captured by the features below. The parse-based analyses, conversely, instantiate a more targeted search.

- AGENT HEAD: word form of the syntactic head of the agent constituent
- AGENT TOPICALITY: *first person, second person, third person pronoun, definite noun, indefinite noun, subordinate clause* (taken over from Pijpops and Speelman 2017: 227–228)
- AGENT COMPLEXITY: Natural logarithm of the number of words of the agent constituent
- VERB FORM: word form of the verb
- THEME HEAD: word form of the syntactic head of the theme constituent
- THEME TOPICALITY: *first person, second person, third person pronoun, definite noun, indefinite noun, subordinate clause* (taken over from Pijpops and Speelman 2017: 227–228)
- THEME COMPLEXITY: Natural logarithm of the number of words of the theme constituent
- THEME-VERB ORDER: *theme-verb, verb-theme*

It would be informative to see which MBL-analyses perform better than the others, but we are not interested in achieving the highest possible accuracy in absolute terms. As such, we perform no parameter search but simply use parameter settings that are regarded as defaults for MBL-analyses, viz. the IB1-algorithm with the overlap metric and gain ratio feature weighting, k set to 5 and inverse linear class voting weights. All presented analyses are the results of leave-one-out-testing (Weiss and Kulikowski 1991; Daelemans et al. 2010: 12, 40).

---

[78] These features were blinded for the choice of variant, for the type of negation and for the choice between pronominal and adverbial realizations of the theme in the prepositional variant (cf. Section 5.1). All words were set in lower case.

## 7.1.3     Testing the procedure

For *verlangen* 'desire', the regression models in Section 6.2 yielded high C-indexes and indicated outspoken lexical biases of the theme. We expect similar results from the MBL-proceure. For *peilen* 'gauge', we also expect a high C-index, given the massive difference between Belgium and the Netherlands (cf. Figure 8 in Section 6.1), and hence also expect a high gain ratio for the feature COUNTRY, and perhaps also some effect for the feature THEME HEAD – though the effect here appeared a lot weaker than for *verlangen* 'desire'. For *slachtoffer zoeken* 'search victim', the MBL-analyses should point toward a distinction of aggressors versus helpers, while for *woord zoeken* 'search word', we expect it to point towards literally looking for specific words versus attempting to express something.

Figure 20 shows the C-indexes of both types of MBL-analyses for *verlangen* 'desire' and *peilen* 'gauge'. We find excellent performance for these verbs, as expected. Figure 21 displays the gain ratios of the analyses. For *verlangen* 'desire', we indeed find that THEME HEAD is the most useful for predicting the variant.

For *peilen* 'gauge', the gain ratio of COUNTRY is – literally – off the charts. We also find high values for VERB-THEME ORDER, VERB FORM and THEME HEAD. For VERB-THEME ORDER, the instances where the theme precedes the verb appear to prefer the transitive variant. This may be explained in terms of complexity, cf. Subsection 4.3.2. The feature VERB FORM points in the same direction: the participles and infinitives have a preference for the transitive variant: these are the verb forms that are always placed in the second verbal pole and where the theme hence always precedes the verb, since we have excluded the instances where the theme is placed in postfield position. Finally, the effect for THEME HEAD was expected.
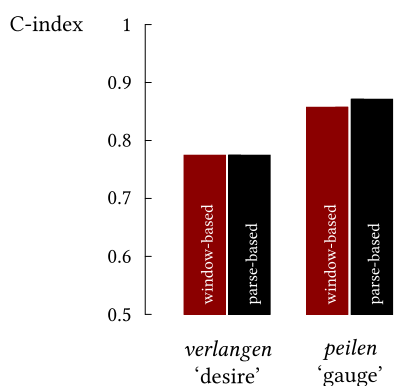


Figure 20: Predictive performance of the MBL-analyses at the verb-level for the verbs *verlangen* 'desire' and *peilen* 'gauge'.

a. **Parse-based** MBL-analyses   b. **Window-based** MBL-analyses
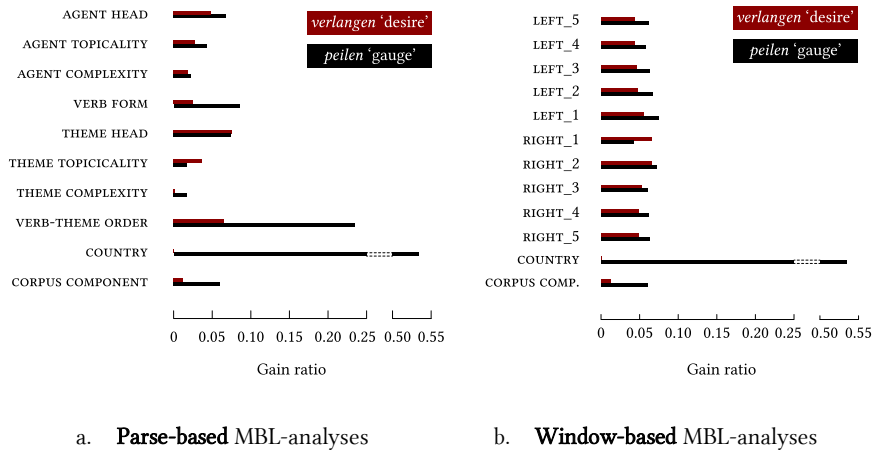
Figure 21: Gain ratios of the MBL-analyses at the verb-level for the verbs *verlangen* 'desire' and *peilen* 'gauge'.

We then turn to *slachtoffer zoeken* 'search victim' and *woord zoeken* 'search word'. Figure 22 and Figure 23 contain their C-indexes and gain ratios. The high C-index for *slachtoffer zoeken* 'search victim' indicates that the analyses managed to identify a strict distinction here. The high gain ratio for AGENT WORD is expected: it shows the distinction between agents acting as aggressors, such as *zakkenrollers* 'pickpockets' or *daders* 'perpetrators', versus agents acting as helpers, such as *reddingswerkers* 'rescue workers' and *duikers* 'divers'. When we look under the hood for COMPONENT, we find that the prepositional variant is much more often used in the auto cues: looking directly at the hypothesis-generating data, it appears that news readers more often talk about large scale disasters where rescue workers are looking for victims than about instances where criminals are searching for victims. As for the feature LEFT_1, i.e. the first word to the left of the place of the preposition, we find *altijd* 'always' or 'still', *brokstuk* 'wreckage', *speurhond* 'tracker dog' and *puin* 'rubble' to indicate the use of the prepositional variant. This again points towards the distinction between aggressors and helpers.

The lower C-index for *woord zoeken* 'search word' signals that the analyses were less successful in picking up on the relevant distinction, or that the distinction itself is less strict. As for the gain ratios, we find an outspoken peak for THEME WORD. This is conspicuous, since this feature does not vary much: it is of course always a form of *woord* 'word'. We indeed find an outspoken difference between these word forms: the singular prefers the transitive variant, while the plural more often appears in the prepositional variant. From this, the relevant distinction could be inferred: literally looking for a word in a text usually involves just one word, while if a speaker means to express some proposition, multiple words are typically needed.
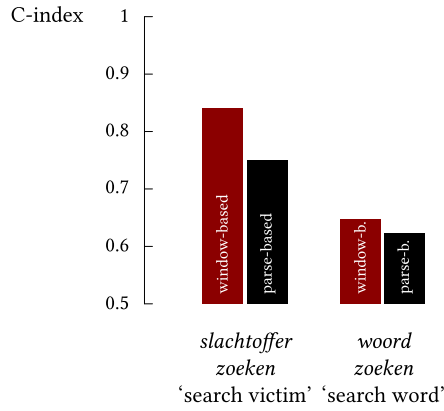
Figure 22: Predictive performance of the MBL-analyses at the object-level for the verb-object combinations *slachtoffer zoeken* 'search victim' and *woord zoeken* 'search word'.
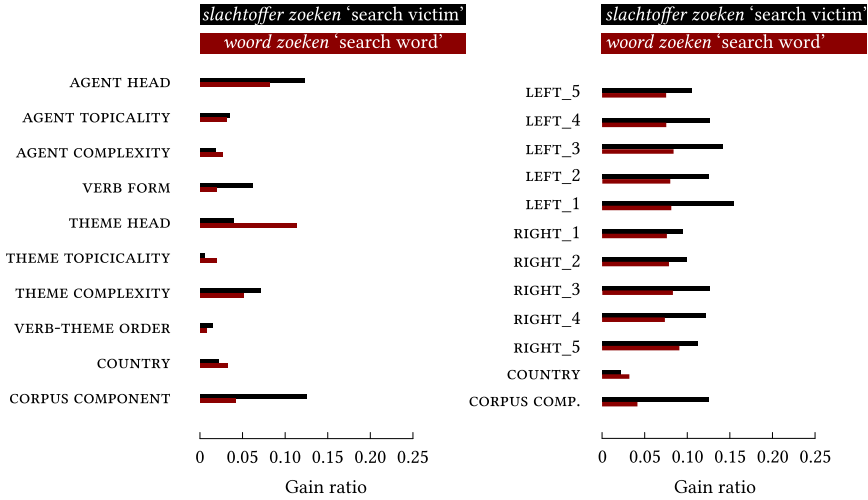


a. **Parse-based** MBL-analyses          b. **Window-based** MBL-analyses

Figure 23: Gain ratios of the MBL-analyses at the object-level for the verb-object combinations *slachtoffer zoeken* 'search victim' and *woord zoeken* 'search word'.

## 7.2    Hypothesis-generation

The MBL-analyses thus seem capable of pointing to the relevant distinctions in data where we know what to find. We now turn to other verbs and verb-theme combinations. We only look at combinations with full nominal themes, i.e. no pronouns or subordinate clauses, and require each of these to minimally yield 40 instances of each variant to be taken up in the analyses. Ideally, this frequency requirement would be set higher, but that would leave hardly any verb-theme combinations other than those with the highly frequent verb *zoeken* 'search'. We then randomly select 30 instances of the transitive variant and 30 instances of the prepositional variant and exclude these from the MBL-analyses, so they can later be used in hypothesis-testing. Next, we run both types of MBL-analyses on the remaining data of each verb or verb-theme combination.

The first subsection discusses the results for the branch of the telephonic verbs *bellen* 'phone' and *telefoneren* 'phone' and the verb-theme combinations *nummer bellen* 'call phone number' and *politie bellen* 'call the police'. These were the only verbs and only verb-theme combinations of this branch that reached the minimum frequency requirement. The 120 instances intended for hypothesis-testing of the combinations *nummer bellen* 'call phone number' and *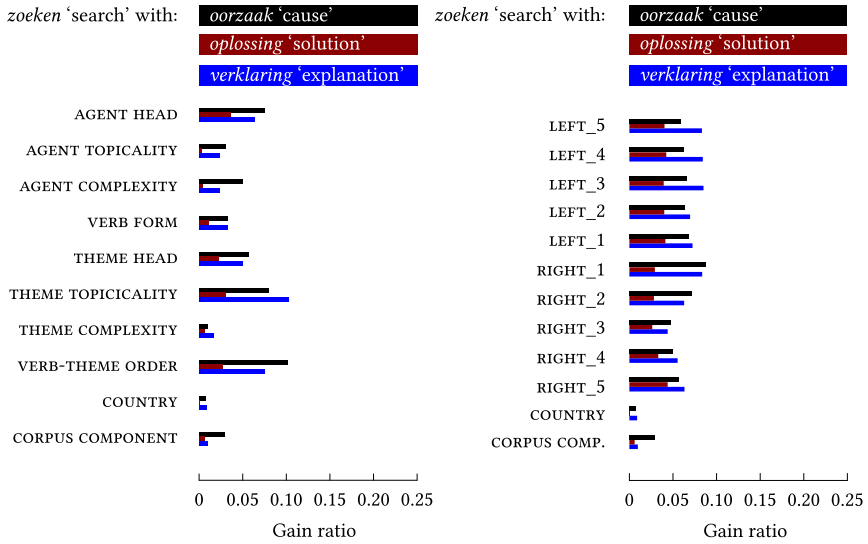politie bellen* 'call the police' were excluded from the data used for hypothesis-generation of the verb *bellen* 'phone'. Conversely, the 60 instances intended for hypothesis-testing of the verb *bellen* 'phone' were not drawn from the instances of *nummer bellen* 'call phone number' and *politie bellen* 'call the police'. In this way, it was ensured that none of the data intended for hypothesis-testing were used in hypothesis-generation.

The second subsection then turns to the branch of the motoric verbs, with the verbs *grijpen* 'grab' and *happen* 'snap' and the verb-theme combination *lucht happen* 'inhale air'. The data for the hypothesis-generation and hypothesis-testing of the verb *happen* 'snap' were selected analogously to the verb *bellen* 'phone' to make sure none of the hypothesis-testing instances of *lucht happen* 'inhale air' were used in the hypothesis-generation procedure of the verb *happen* 'snap' and vice versa.

Finally, the third subsection looks at the verb-theme combinations with the highly frequent verb *zoeken* 'search'. For *zoeken* 'search', we have 24 combinations that reach the frequency requirement – which is a lot to report on. We will hence only look at the combinations where the C-index of the parse-based or window-based analysis is at least 0.6. The lower the C-index, the more unlikely it is that the analyses have picked up on a relevant distinction. This leaves us with 12 combinations, viz. *zoeken* 'search' with *alternatief* 'alternative', *antwoord* 'answer', *dader* 'perpetrator', *manier* 'manner', *mogelijkheid* 'possibility', *oorzaak* 'cause', *oplossing* 'solution', *spoor* 'track', *verklaring* 'explanation', *vorm* 'shape', *waarheid* 'truth' and *weg* 'way'.

## 7.2.1    Telephonic verbs

The C-indexes and the gain ratios for *bellen* 'phone', *telefoneren* 'phone', *nummer bellen* 'phone number' and *politie bellen* 'phone police' can be found respectively in Figure 24 and Figure 25.

The gain ratios of *bellen* 'phone' and *telefoneren* 'phone' seem to follow a similar pattern, with high values for the parse-based features relating to the theme, most notably THEME HEAD. When we look under the hood at the themes in the hypothesis-generating datasets, it appears that human themes more often occur in the transitive variant, while non-human themes, i.e. collectives and inanimates, seem to prefer the prepositional variant. In the window-based analyses, we also find a little peak at RIGHT_1, i.e. the first word on the right side of where the preposition would be placed. This is the position where either the theme head or the determiner of the theme is placed: this feature appears to point in the same direction of human vs. non-human themes.

Finally, the feature COUNTRY yields a relatively high gain ratio for *bellen* 'phone'. The transitive variant seems more popular in the Netherlands than in Belgium, which is not unexpected (cf. Figure 8 in Section 6.1). We then formulate the following hypothesis for *bellen* 'phone' and *telefoneren* 'phone': when the theme is human, the transitive variant will be preferred, whereas when the theme is not human, the prepositional variant will be more likely chosen.
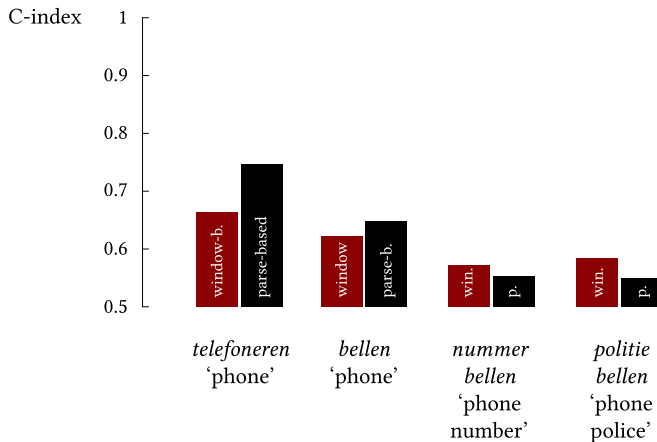


Figure 24: Predictive performance of the MBL-analyses for the verbs *telefoneren* 'phone' and *bellen* 'phone' and the verb-object combinations *nummer bellen* 'phone number' and *politie bellen* 'phone police'.

a.  **Parse-based** MBL-analyses          b.  **Window-based** MBL-analyses
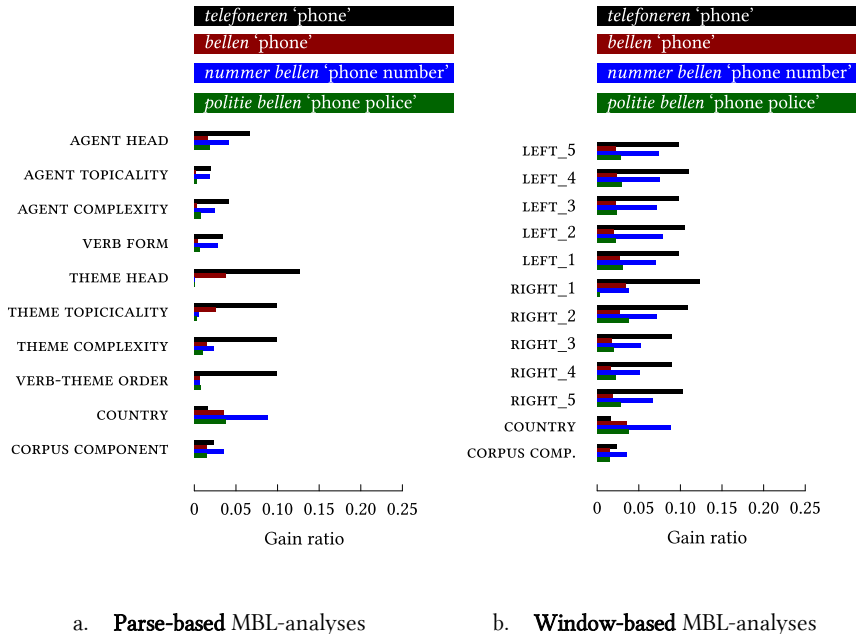
Figure 25: Gain ratios of the MBL-analyses for the verbs *telefoneren* 'phone' and *bellen* 'phone' and the verb-object combinations *nummer bellen* 'phone number' and *politie bellen* 'phone police'.

The MBL analyses for *nummer bellen* 'call phone number' and *politie bellen* 'call the police' yield low C-indexes, indicating that they are unlikely to have detected a useful distinction. For *nummer bellen* 'call phone number', we again find a peak for COUNTRY, and also for AGENT HEAD and RIGHT_2, at least relative to the other words on the right. For AGENT HEAD, we find that the subordinate clauses with enclosed antecedents as in (182a-b) tend to occur with the prepositional variant, while pronouns as in (182c-d) prefer the transitive variant. For RIGHT_2, we find that words like *centrale* 'central', *vast* 'land' as in 'land line', and *speciaal* 'special' are indicative of the prepositional variant, while *nummer* 'number', *nummers* 'number' and *eigen* 'own' indicate the transitive variant. Based on these indications, and the results for *bellen* 'phone' and *telefoneren* 'phone', we formulate the hypothesis that when the owner of the number is human, transitive variant will be preferred, whereas when the owner of the number is not human, the prepositional variant will be more probable.

182. a.    *Wie informatie heeft over   de fiets, kan  het gratis  nummer van*
        who  information has   about  the bike  can  the free   number  of
        *de    politie   bellen.*
        the   police    call
        'Who has information about the bike, can call the free number of the
        police.'
                (WS-U-E-A-0000052456.p.1.s.1)

    b.    *Wie er    bij   wil    zijn, belt     naar het nummer 050/71.24.01.*
        who  there with  wants be    phones  to    the number 050/72.24.01
        'Who wants to be there, should call the number 050/72.24.01.'
                (WR-P-P-G-0000384907.p.2.s.2)

    c.    *Eerst belde    ze het nummer van Jones, en   toen   hij niet*
        first   phoned  she the number of    Jones  and  when  he  not
        *opnam  sprak  ze een  boodschap in   op zijn voicemail.*
        took_up  spoke  she a      message    PART on his   voicemail
        'First she called the number of Jones, and when he did not answer, she
        left a message on his voicemail.'
                (WR-P-P-B-0000000229.p.2056.s.1)

    d.    *Ik bel     direct       naar het nummer in  het display  van  mijn*
        I  phone  immediately to    the number on  the display  of    my
        *mobiel.*
        mobile_phone
        'I immediately call the number on the display of my mobile phone.'
                (WR-P-P-B-0000000242.p.575.s.2)

Of course, this distinction cannot be helpful for *politie bellen* 'phone police', since *politie* 'police' is always collective. The C-index of the parse-based analysis for this combination is really low, and its gain ratios do not seem to yield anything useful. Its window-based analysis performs somewhat better – though still bad. Here, we find a little peak for the gain ratio of RIGHT_2. This is notable, since this feature is almost always *politie* 'police'. When it is, the transitive variant appears more likely, while when it is instead an adjective like *lokale* 'local', the prepositional variant seems preferred. Perhaps when one calls the police in an emergency, it does not matter which type of police is called, and then the transitive variant is preferred. Meanwhile, when one calls to police to report something non-urgent, it is more likely that the type of police is specified by an adjective, such as *lokale* 'local', *federale* 'federal', *echte* 'real', etc. In that case, it could be conjectured that the prepositional variant is preferred. This then forms our hypothesis for *politie bellen* 'call the police'.

## 7.2.2    Motoric verbs

Figure 26 and Figure 27 contain the C-indexes and gain ratios of the MBL-analyses at the level of the verb for *grijpen* 'grab' and *happen* 'snap', and at the level of the object for *lucht happen* 'inhale air'. For *grijpen* 'grab', whose analyses reach reasonable C-indexes, we find peaks in gain ratio for the features THEME HEAD, AGENT HEAD and RIGHT_2. The features THEME HEAD and RIGHT_2 seem to indicate the same information: when the object being grabbed is *macht* 'power', *leiding* 'lead', *titel* 'title' or the like, the transitive variant is preferred, while when it is *wapens* 'weapons', *mes* 'knife', *telefoon* 'telephone', etc., the prepositional variant is more often chosen. For AGENT HEAD, we find that the transitive variant tends to be used when the army, a general, or the military are grabbing something, while when a personal pronoun is grabbing something, the prepositional variant occurs more often. We then hypothesize that when *grijpen* can be translated as 'conquer', the transitive variant will be used, whereas when it involves a grabbing of a concrete object to be immediately used, the prepositional variant will be preferred.

For the verb *happen* 'snap', Figure 27 show peaks at THEME COMPLEXITY, THEME HEAD and RIGHT_1. We find that most themes are only a single word long, but when they are longer, the transitive variant is much more likely. This is notable, since it runs counter to most of our expectations based on language complexity (cf. Section 4.3). When the THEME HEAD is *lucht* 'air', the transitive variant is preferred, versus the prepositional variant when it is *adem* 'breath'. The feature RIGHT_1 appears to contain the same information.
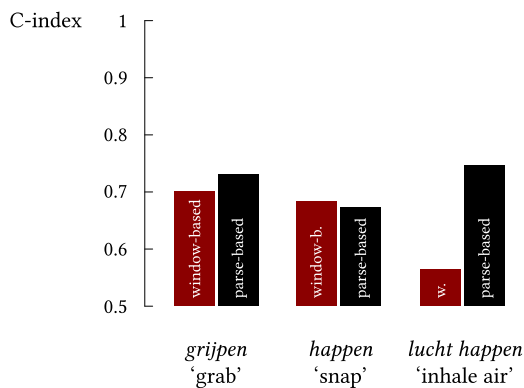


Figure 26: Predictive performance of the MBL-analyses for the verbs *grijpen* 'grab' and *happen* 'snap' and the verb-theme combination *lucht happen* 'inhale air'.

a.  **Parse-based** MBL-analyses          b.  **Window-based** MBL-analyses
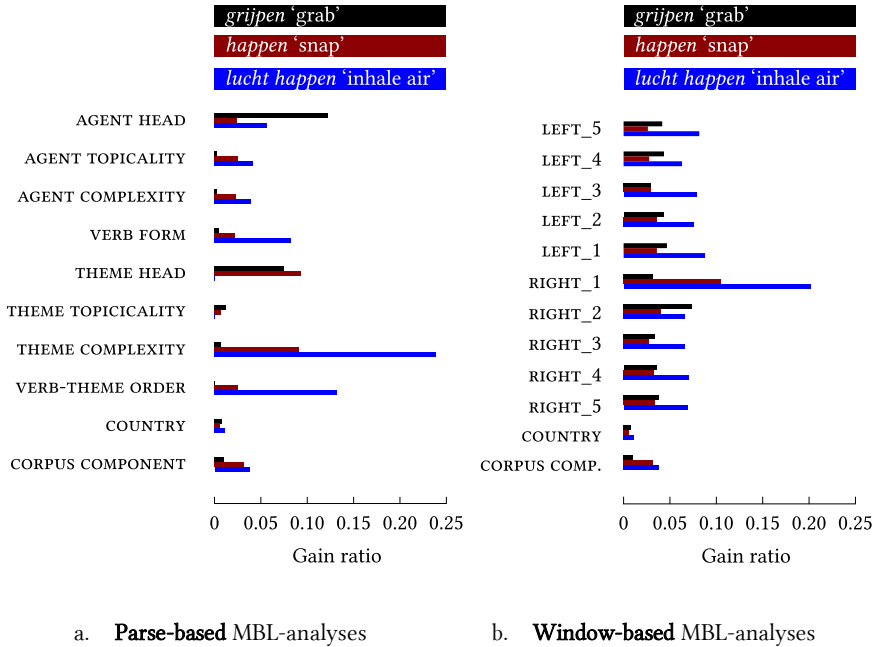
Figure 27: Gain ratios of the MBL-analyses for the verbs *grijpen* 'grab' and *happen* 'snap' and the verb-theme combination *lucht happen* 'inhale air'.

For the verb-theme combination *lucht happen* 'inhale air', the massive peak at THEME COMPLEXITY signals the same tendency as for *happen* 'snap'. The outspoken peak at RIGHT_1 is also noteworthy, especially since this feature is almost always *lucht* 'air'. When it is not, however, but rather *wat* 'some', *verse* 'fresh' or *frisse* 'fresh', then this is a clear tell-tale for the transitive variant.

These indications could be interpreted as follows. When the inhaling clearly succeeds – and it can hence be specified by adjectives what sort of air is inhaled – the transitive variant is used. Conversely, when it is unclear whether the inhaling succeeds, for instance when someone is drowning or choking, then (i) the theme *adem* 'breath' will be more likely used than *lucht* 'air', and (ii) when *lucht* 'air' is nonetheless used, it does not matter whether the air in question is fresh or not. In that case, the prepositional variant would be used. This is thus a hypothesis in terms of conation. We will also test the same distinction for *grijpen* 'grab' (cf. Section 2.3).

## 7.2.3     Combinations with *zoeken* 'search'

Figure 28 contains the C-indexes of all analyses for specific themes of the verb *zoeken* 'search' where one of the analyses reached at least a C-index of 0.6. We begin with 6 themes where the gain ratios led us to formulate the same hypothesis, viz. *alternatief* 'alternative', *antwoord* 'answer', *mogelijkheid* 'possibility', *oorzaak* 'cause', *oplossing* 'solution' and *verklaring* 'explanation'. These gain ratios are shown in Figure 29 and Figure 30.
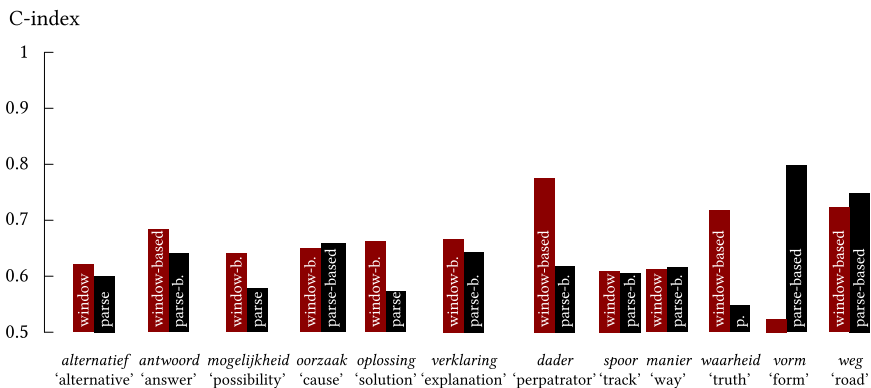


Figure 28: Predictive performance of the MBL-analyses for the selected verb-theme combinations of the verb *zoeken* 'search'.

Here, the peaks for THEME TOPICALITY and THEME HEAD indicate that definite and singular themes trigger the transitive variant, while indefinite and plural themes ones trigger the prepositional variant. We find the same information in the peaks of RIGHT_1 for *oorzaak* 'cause' and *verklaring* 'explanation'. Based on this, we speculate that if the alternative, cause, explanation, etc. is already more or less known, but merely needs to be acquired or made into reality, then the transitive variant is employed, whereas if that is not the case, but an actual search still needs to be carried out, the prepositional variant will be preferred. To operationalize this distinction, we will distinguish between those instances where a locative adjunct already marks where the alternative, solution etc. needs to be sought, as in (183), versus instances where a locative adjunct is not present, as in (184). The first group is predicted to prefer the transitive variant, while the second group would prefer the prepositional variant.
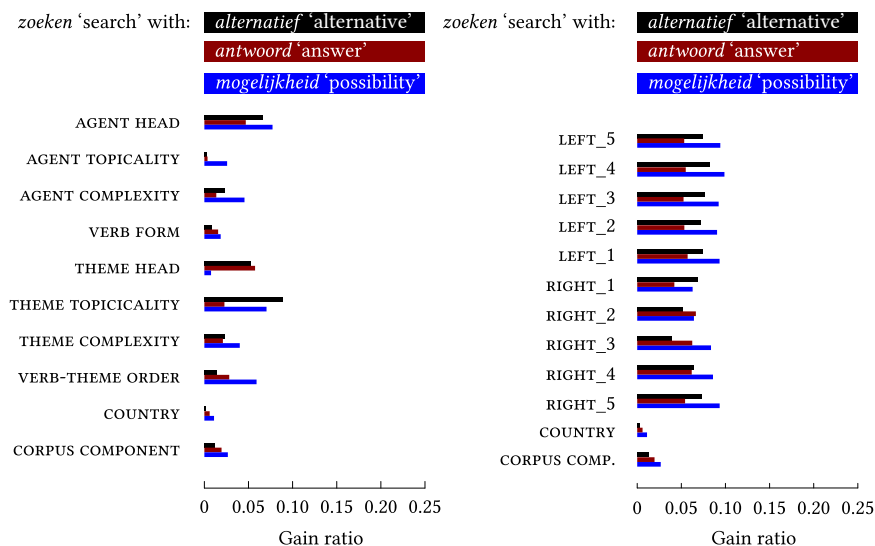
a.   **Parse-based** MBL-analyses     b.   **Window-based** MBL-analyses

Figure 29: Gain ratios of the MBL-analyses for the verb-theme combinations *alternatief zoeken* 'search alternative' and *antwoord zoeken* 'search answer' and *mogelijkheid zoeken* 'search possibility'.

It is not immediately clear what the peaks for AGENT HEAD indicate. For *oorzaak* 'cause', agents like *experts* 'experts', *politie* 'police' and *parket* 'public prosecutor' seem to prefer the prepositional variant, which may point towards the same distinction. In general, in occurrences where the agent is not expressed and that are labelled as such – e.g. passives, imperatives, infinitival clauses such as (185a) or cases of ellipsis such as (185b) – the transitive variant seems to be preferred, although for *mogelijkheid zoeken* 'search possibility', the prepositional variant is actually more frequent in such instances.

Figure 30: Gain ratios of the MBL-analyses for the verb-theme combinations *oorzaak zoeken* 'search cause', *oplossing zoeken* 'search solution' and *verklaring zoeken* 'search explanation'.

183. a. *Oliebedrijven  zoeken  daarom    alternatieven in  de Atlantische*
        Oil_companies  search    therefore  alternatives   in   the Atlantic
        *Oceaan, voor    de kust    van  West-Afrika.*
        Ocean    before  the coast  of     West-Africa
        'Oil companies are therefore looking for alternatives in the Atlantic
        Ociean, in front of the coast of West-Africa.'
              (WR-P-P-G-0000048286.p.1.s.3)

     b. *Veel    zoeken   nu    naar goedkopere alternatieven in  het*
        many     search   now   to    cheaper      alternatives   in   the
        *binnenland.*
        interior
        'Many now look for cheaper alternatives in the interior.'
              (WR-P-P-G-0000676377.p.2.s.6)

    c.     *Betcke zoekt    de oplossing in beter   onderwijs (…)*
         Betcke searches the solution    in   better   education
         'Betcke searches the solution in better education (…)'
                 (WR-P-P-G-0000151054.p.5.s.1)

    d.     *De   club zoekt    in eigen club naar een   oplossing om de*
         The club searches in own    club to    a     solution    to the
         *zwaar    gekwetste Harald Pinxten te   vervangen.*
         heavily injured     Harald Pinxten to   replace
         'The club is now searching in its own club for a solutions to replace the
         heavily injured Harald Pinxten.'
                 (WR-P-P-G-0000664719.p.9.s.2)

184. a.     *We     moeten alternatieven      zoeken.*
         we      must     alternatives      search
                 (WR-P-P-G-0000098226.p.27.s.5)

    b.     *De directie    zoekt    nog   naar alternatieven, (…)*
         the direction searches still    to     alternatives
         'The direction is still searching alternatives.'
                 (WS-U-E-A-0000351428.p.1.s.4)

    c.     *En     toch     zoekt   niemand nu      een      oplossing.*
         and     still     searches no_one   now      a       solution
         'And still no one is looking for a solution now.'
                 (WR-P-P-H-0000110728.p.18.s.3)

    d.     *Nijmegen zoekt    naar een   oplossing.*
         Nijmegen searches to     a      solution
         'Nijmegen is searching a solution.'
                 (WR-P-P-G-0000083223.p.8.s.1)

185. a.     *Het ligt   in   de menselijke aard   om positief te   zijn en     steeds*
         It      lies in the human       nature to positive to be    and always
         *weer   nieuwe   mogelijkheden te   zoeken, (…)*
         again new      possibilities     to search
         'It is part of human nature to be positive and always look for new
         possibilities.'
                 (WR-P-P-H-0000021391.p.5.s.1)

b.  *Zocht te vaak naar de moeilijke oplossing, terwijl een simpele*
    searched too often to the hard solution while a simple
    *pass over een paar meter werd gevraagd.*
    pass over a few meter was asked
    'Searched too often for a difficult solution, when a simple pass over a
    few meters was asked.'
                    (WR-P-P-G-0000030999.p.21.s.2)

Figure 31 contains the gain ratios for *dader zoeken* 'search perpetrator' and *spoor zoeken* 'search track'. Both, especially *dader zoeken* 'search perpetrator', score high gain ratios for the features relating to the agent. Here, we find *politie* 'police', *speurders* 'investigators', *recherche* 'criminal investigation department' etc. to trigger the prepositional variant. For the feature LEFT_1, we find *nog* 'still', *steeds* 'still', *intensief* 'intensively' and again *politie* 'politie' to indicate the prepositional variant. For the feature RIGHT_1 of *dader zoeken* 'search perpetrator', the level *dader* 'perpetrator' is a reliable indicator of the transitive variant – which is not surprising, since these instances are all headlines. For RIGHT_1 of *spoor zoeken* 'search track', we find *sporen* 'tracks' or 'leads', *nieuwe* 'new' and *bruikbare* 'usable' to prefer the prepositional variant. We interpret these indications to point towards a distinction between sentences in which the authorities search for leads in a police investigation, which we expect to prefer the prepositional variant, versus other forms of searching tracks, which would prefer the transitive variant.



a.  **Parse-based** MBL-analyses        b.  **Window-based** MBL-analyses
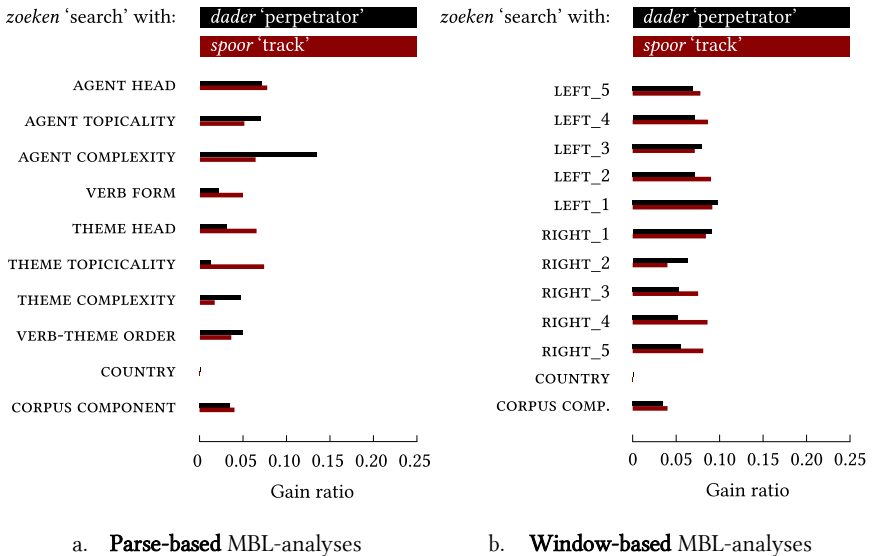
Figure 31: Gain ratios of the MBL-analyses of the verb-theme combinations of the verb *dader zoeken* 'search perpetrator' and *spoor zoeken* 'search track'.

Finally, Figure 32 shows gain ratios for the combinations of *zoeken* 'search' with *manier* 'way', *waarheid* 'truth', *vorm* 'shape', and *weg* 'road'. For *manier zoeken* 'search way' and *waarheid zoeken* 'search truth' we do not readily find a potentially relevant distinction. For *manier zoeken* 'search way', the feature VERB-THEME ORDER indicates that the prepositional variant is more frequent when the verb precedes the theme, which is a stable effect that we also find in the other analyses. The feature THEME TOPICALITY indicates that the prepositional variant is more popular among indefinite themes, which actually goes against the results for *alternatief* 'alternative', *mogelijkheid* 'cause', etc. The features of the window-based analyses also do not seem to yield anything easily interpretable.

For *waarheid zoeken* 'search truth', only the window-based analysis performs well. The high gain ratios for LEFT_5, LEFT_4 and LEFT_3 seem to indicate that the transitive variant is preferred near the beginning of sentences, when these features have value *x*. Other than that, the peaks do not seem readily interpretable. As a result, since we cannot come up with anything better, we will annotate these combinations for the presence of a locative adjunct, in analogy to *alternatief* 'alternative', *mogelijkheid* 'cause', etc.
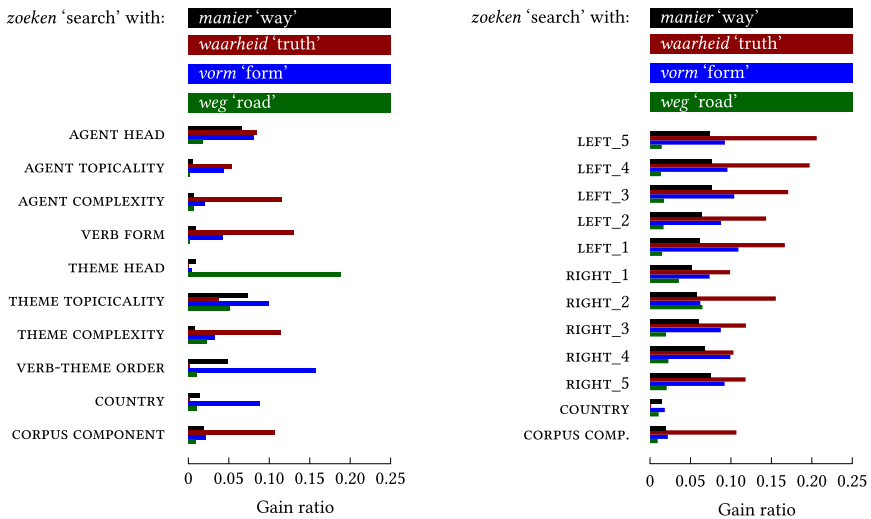


a.   **Parse-based** MBL-analyses          b.   **Window-based** MBL-analyses

Figure 32: Gain ratios of the MBL-analyses of the verb-theme combinations *manier zoeken* 'search way', *waarheid zoeken* 'search truth', *vorm zoeken* 'search shape' and *weg zoeken* 'search road'.

For *vorm* 'shape', Figure 28 indicates an excellent performance of the parse-based analysis, with peaks in gain ratio for VERB-THEME ORDER, THEME TOPICALITY and AGENT HEAD. VERB-THEME ORDER shows its usual effect, viz. a preference for the prepositional variant when the verb precedes the theme. Looking under the hood of THEME TOPICALITY and AGENT HEAD, we seem to find a distinction between sportspeople trying to get into their best condition, which triggers the prepositional variant, and other instances of searching for forms.

Lastly, for *weg* 'road', both the parse-based and window-based analysis perform well (cf. Figure 28). The window-based analysis exhibits an outspoken peak in gain ratio for RIGHT_2 and a smaller one for RIGHT_1. For RIGHT_2, the levels *eigen* 'own' and *weg* 'road' trigger a choice for the prepositional variant, while *wegen* 'roads' or 'ways' and *om* 'for' prefer the prepositional variant. For RIGHT_1, we find possessive pronouns to be indicative of a choice for the transitive variant, whereas *wegen* 'roads' and *nieuwe* 'new' prefer the prepositional variant. The peak for THEME HEAD is also noteworthy, since this feature hardly varies: it is either singular *weg* 'road' or plural *wegen* 'roads': we again find the singular to trigger the transitive variant, and the plural the prepositional variant. Based on this, we formulate the hypothesis that when someone is looking for their place in society or in a new job position or the like, this will be expressed in the transitive variant, while other forms of *weg zoeken* 'search road', e.g. looking for ways to do something, will more often be expressed in the prepositional variant

## 7.3    Hypothesis-testing

The 30 randomly selected transitive and prepositional instances of each alternation that had been kept separate were then annotated for the distinctions hypothesized in the previous section. During the manual annotation, the instances were again blinded for their choice of variant to prevent any bias on the part of the annotator. Figure 33 contains the mosaic plots showing the results of this annotation. Table 20 presents an overview of these results.
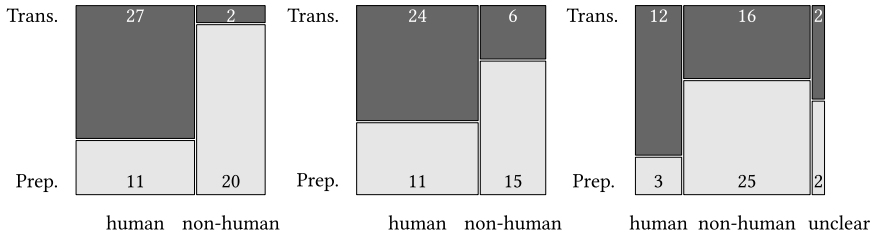
a.   ADDRESSEE ANIMACY for *telefoneren* 'phone' (Chisq. test: p < 0.0001, Cramer's V = 0.60)

b.   ADDRESSEE ANIMACY for *bellen* 'phone' (Chisq. test: p = 0.0149, Cramer's V = 0.31)

c.   ADRESSEE ANIMACY for *nummer bellen* 'phone number (*unclear* excluded, Chisq. test: p = 0.0066, Cramer's V = 0.36)
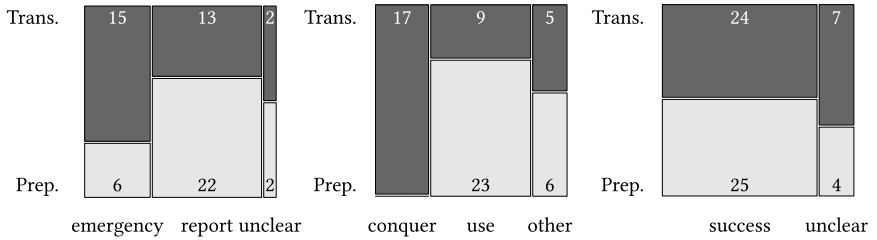


d.   TYPE OF CALL for *politie bellen* 'phone police' (*unclear* excluded, Chisq. test: p = 0.013, Cramer's V = 0.33)

e.   MEANING GRIJPEN for *grijpen* 'grab' (*other* excluded, Chisq. test,: p < 0.0001, Cramer's V = 0.69)

f.   CONATION for *grijpen* 'grab' (Chisq. test: p = 0.3794, Cramer's V = 0.11)
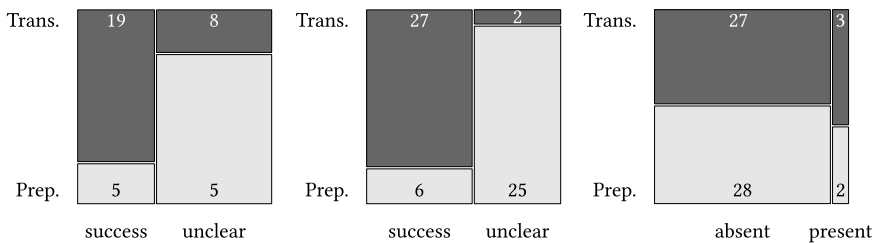


g.   CONATION for *happen* 'snap' (Chisq. test: p < 0.0001, Cramer's V = 0.56)

h.   CONATION for *lucht happen* 'inhale air' (Chisq. test: < 0.0001, Cramer's V = 0.74)

i.   LOCATIVE ADJUNCT for *Alternatief zoeken* 'search alternative' (Fisher's exact test: p > 0.999, Cramer's V = 0.06)
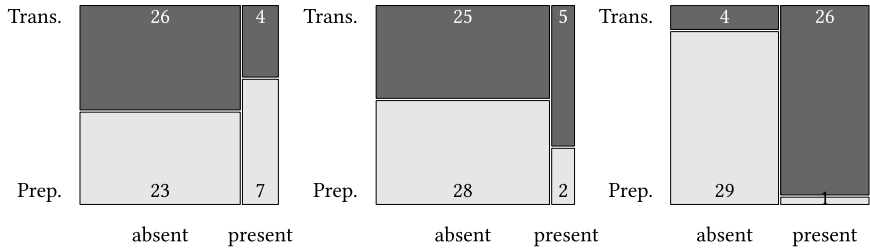
j. LOCATIVE ADJUNCT for *antwoord zoeken* 'search answer' (Chisq. test: p = 0.3169, Cramer's V = 0.13)

k. LOCATIVE ADJUNCT for *mogelijkheid zoeken* 'search possibility' (Fisher's exact test: p = 0.4238, Cramer's V = 0.16)

l. LOCATIVE ADJUNCT for *oorzaak zoeken* 'search cause' (Fisher's exact test: p < 0.0001, Cramer's V = 0.84)
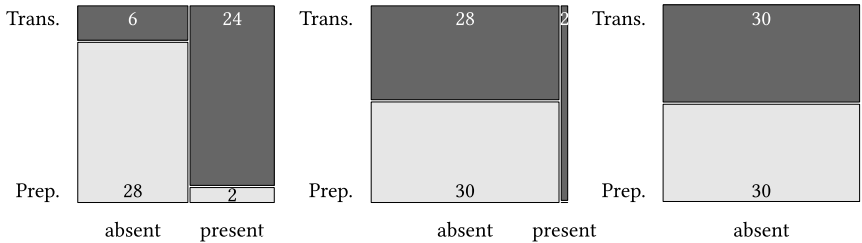
m. LOCATIVE ADJUNCT for *verklaring zoeken* 'search explanation' (Chisq. test: p < 0.0001, Cramer's V = 0.74)

n. LOCATIVE ADJUNCT for *oplossing zoeken* 'search solution' (Fisher's exact test: p = 0.4915, Cramer's V = 0.19)

o. LOCATIVE ADJUNCT for *manier zoeken* 'search way'
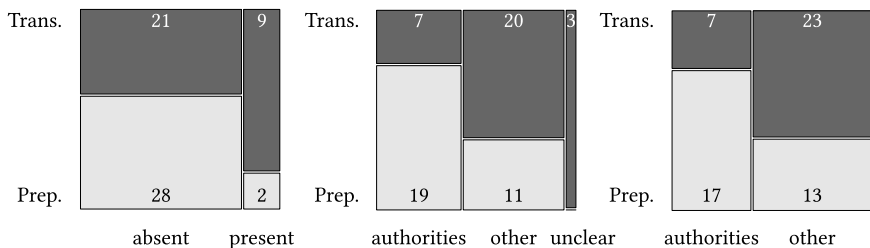
p. LOCATIVE ADJUNCT for *waarheid zoeken* 'search truth' (Chisq. test: p = 0.0195, Cramer's V = 0.30)

q. AGENT TYPE for *dader zoeken* 'search perpetrator' (unclear excluded, Fisher's exact test: p = 0.0028, Cramer's V = 0.43)

r. AGENT TYPE for *spoor zoeken* 'search track' (Chisq. test: p = 0.0084, Cramer's V = 0.34)

Trans. 3 | 27   Trans. 13 | 17

Prep. 15 | 15   Prep. | 30

sport | other   one's place | other
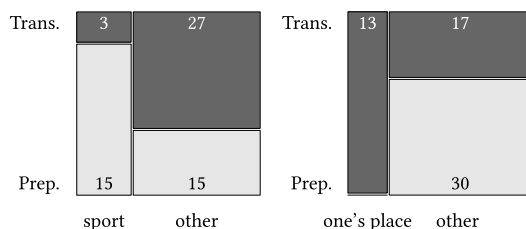
s.  MEANING VORM for *vorm zoeken* 'search form' (Chisq. test: p = 0.0007, Cramer's V = 0.44)

t.  MEANING WEG for *weg zoeken* 'search road' (Chisq. test: p < 0.0001, Cramer's V = 0.53)

Figure 33: Mosaic plots of the hypothesis testing analyses.

| Verbal group | Variable | Hypothesis and levels | | Verb or verb-theme combination | | Result |
|---|---|---|---|---|---|---|
| Telephonic verbs | ADDRESSEE ANIMACY | Trans.: Prep: Rest cat.: | *human* *non-human* *unclear* | *bellen* *telefoneren* *nummer bellen* | 'phone' 'phone' 'phone number' | confirmed confirmed confirmed |
| | TYPE OF CALL | Trans.: Prep.: Rest cat.: | *emergency* *report* *unclear* | *politie bellen* 'phone police' | | confirmed |
| Motoric verbs | MEANING GRIJPEN | Trans.: Prep: Rest cat.: | *conquer* *use* *other* | *grijpen* | *'grab'* | confirmed |
| | CONATION | Trans: Prep: | *success* *unclear* | *grijpen* *happen* *lucht happen* | 'grab' 'snap' 'inhale air' | not confirmed confirmed confirmed |
| *zoeken* 'search' | LOCATIVE ADJUNCT | Trans: Prep: | *present* *absent* | *alternatief* *antwoord* *mogelijkheid* *oorzaak* *verklaring* *oplossing* *manier* *waarheid* | 'alternative' 'answer' 'possibility' 'cause' 'explanation' 'solution' 'way' 'truth' | not confirmed not confirmed not confirmed confirmed confirmed not confirmed not confirmed confirmed |
| | AGENT TYPE | Trans.: Prep.: | *other* *authorities* | *dader* *spoor* | 'perpetrator' 'track' | confirmed confirmed |
| | MEANING VORM | Trans.: Prep.: | *other* *sport* | *vorm* | 'form' | confirmed |
| | MEANING WEG | Trans.: Prep.: | *one's place* *other* | *weg* | 'road' | confirmed |

Table 20: overview of the results of the hypothesis testing analyses

For the telephonic verbs, the distinction between human and non-human addressees appears successful in predicting the choice of variant. This could be interpreted as a distinction in terms of directionality – although it should be stressed that this is a post-hoc interpretation and that it seriously stretches the notion of directionality. When we call other humans, we expect a two-way interaction with them, with information flowing in both directions. Conversely, when calling companies or services, we are typically calling to report something, which involves more of a one-directional information flow. When calling inanimate telephone machines, there is no information exchange whatsoever, merely a telephone signal being sent from one place to another. The latter two thus involve more of a one-directional action, which could be argued to trigger the prepositional variant.

This cannot be the whole story, however. The transitive variant also seems to be used when calling someone or a service to come over. It was already noted in Subsection 5.2.10.1 that ordering e.g. a taxi to come over would be expressed in the transitive variant. Calling the police to come over in cases of emergency also appears to be more often expressed in the transitive variant. Conversely, merely reporting information to the police seems to be preferably expressed in the prepositional variant, in accordance with the distinction mentioned above.

Turning to the motoric verbs, we find a clear distinction in terms of conation for *happen* 'snap' at the verb level, and for the combination of this verb with *lucht* 'air'. For *grijpen* 'grab', however, we have failed to confirm such a distinction in a straightforward way. Still, it should be noted that instances where the act of grabbing clearly failed were already excluded from the data (cf. Subsection 5.2.5.1), since they did not allow the individual language user to opt for the transitive variant. We could hence only distinguish between those instances where success was immediately apparent or those where it was not.

The distinction between *grijpen* meaning 'conquer' and *grijpen* meaning 'use' was confirmed for *grijpen* 'grab'. This distinction may actually also be interpreted in terms of conation. In cases of grabbing power or grabbing victory, as in (186), it is very possible that the grabbing might fail, and it is hence crucial to distinguish between successful and unsuccessful instances. As such, the use of the prepositional variant may be strictly reserved for unsuccessful instances. By contrast, for instances where concrete objects are being grabbed with the intention to use them, as in (187), there is rarely any doubt that the grabbing will succeed. As a result, the prepositional variant may also be used for instances where the grabbing is indeed succesful, as in (187). Still, this distinction in terms of conation does leave a lot of unexplained variation among instances like (187) or (188) that involve a clearly successful grabbing of a concrete object. It is currently unclear to us what may determine to presence of *naar* in these instances.

186. a.   *De volgelingen van Osman grepen opnieuw de macht en*
          the followers   of   Osman grabbed again    the power and
          *maakten het kalifaat   erfelijk.*
          made      the caliphate  hereditary
          'The followers of Osman again grabbed power and made the caliphate
          hereditary.'
                    (WR-P-P-H-0000171447.p.1.s.5)

     b.   *Roel Kinable, die   een perfo  greep    tegen   Simon De Smet (6-4,*
          Roel Kinable who  a    perfo  grabbed against  Simon De Smet (6-4,
          *6-4), stond  voor een  zware taak tegen    Mario Mertens (...)*
          6-4)  stood  for   a     heavy task against  Mario Mertens
          'Roel Kinable, who grabbed a perfo against Simon De Smet (6-4, 6-4),
          faced a serious challenge against Mario Mertens.'
                    (WR-P-P-G-0000292988.p.2.s.2)

187. *(...)waarop zij naar een  mes greep    en   de man in de rug stak.*
     whereon she to   a    knife grabbed and  the man in  the back stabbed
     '(...) whereupon she grabbed a knife and stabbed the man in his back.'
                    (WR-P-P-G-0000633533.p.2.s.2)

188. a.   *(...)dat de man een keukenmes    greep    en   het koppel*
          that the man a    kitchen_knife grabbed and   the couple
          *neerstak.*
          down_stabbed
          '(...) that the man grabbed a kitchen knife and stabbed the couple
          down.'
                    (WS-U-E-A-0000342304.p.1.s.3)

     b.   *Na    een onbenullige woordenwisseling greep    G. een    mes en*
          after a    trivial      argument         grabbed  G. a     knife and
          *stak    Gregory in de borst.*
          stabbed  Gregory in  the chest
          'After a trivial argument, G. grabbed a knife and stabbed Gregory in
          the chest.'
                    (WR-P-P-G-0000286787.p.3.s.3)

     c.   *De man greep    ook een keukenmes    en   bedreigde  haar*
          the man grabbed also a    kitchen_knife and  threatened her
          *daarmee.*
          therewith
          'The man also grabbed a kitchen knife and threatened her with it.'
                    (WR-P-P-G-0000347566.p.2.s.6)

d.  *Op een  bepaald ogenblik greep    de man naar een  mes en*
    on a    certain  moment grabbed the man to    a      knife and
    *stak    zijn  vrouw verschillende malen.*
    stabbed his  wife    multiple      times
    'On a certain moment, the man grabbed a knife and stabbed his wife
    multiple times.'
    (WR-P-P-G-0000239364.p.3.s.2)

e.  *De man greep    naar een  mes en    stak    een van zijn*
    the man grabbed to    a      knife and stabbed one of   his
    *kameraden  twee keer.*
    comrades    two  times
    'The man grabbed a knife and stabbed one of his comrades twice.'
    (WR-P-P-G-0000245362.p.2.s.4)

f.  *Daarop greep   de dief naar zijn mes en   bedreigde  er   de*
    thereon grabbed the thief to   his  knife and threatened there the
    *boswachter mee.*
    forester    with
    'Thereupon, the thief grabbed his knife and threatened the forester
    with it.'
    (WR-P-P-G-0000264001.p.3.s.4)

Among the combinations of specific themes with the verb *zoeken* 'search', the
distinction between instances with a locative adjunct and those without does not
seem very successful at first sight: it is more often not confirmed than it is
confirmed. There are still reasons to take this distinction seriously, however. First,
our failures to confirm the distinction are possibly due to the lack of instances with
a locative adjunct in the hypothesis-testing data. For *oplossing* 'solution', all
instances with a locative adjunct actually exhibit the transitive variant, as
predicted, but our hypothesis-testing data only happen to contain two such
instances. For *alternatief* 'alternative', *mogelijkheid* 'possibility' and *manier* 'way',
the instances with locative adjuncts are also too infrequent to confirm or refute the
distinction. Second, when the distinction is confirmed, it yields strong effects. For
*oorzaak zoeken* 'search cause', the distinction yields a Cramer's V of 0.84, and for
*verklaring zoeken* 'search explanation', it yields one of 0.74. The effect is
admittedly weaker for *waarheid zoeken* 'search truth', with a Cramer's V of 0.30.

   If we accept that this distinction is relevant for at least some themes of the verb
*zoeken* 'search', then it is still not entirely clear how it should be interpreted,
however. One possibility is to view it as a semantic distinction, as proposed when
formulating the hypothesis. This would correspond to the distinction between 'seek
to make/acquire' and a more literal 'look for' (cf. Subsection 6.2.3): when the cause

or the explanation is already known, as in (189a-b), and hence no actual searching needs to be carried out, this would be expressed in the transitive variant, whereas when that is not the case, as in (189c-d), this would trigger the use of the prepositional variant.

However, it can also be argued that the distinction is not caused by semantics, but rather by processing pressures. In sentences such as (189a-b), the crucial information is not contained in the theme *oorzaak* 'cause' or *verklaring* 'explanation', but rather in the locative adjunct. Hence, the theme would be low in information content and would therefore not be introduced by a preposition (cf. Subsection 4.3.1.2). Yet another possibility is of course that the distinction was originally caused by differences in information density, but that it has already been reinterpreted as a semantic distinction by some language users, e.g. following the lexical-origin mechanism as proposed in Figure 4 of Subsection 4.2.2.2. We will return to this in Section 9.1.

189. a.　*Visserman zoekt　de oorzaak zelf bij　de 'verwevenheid met*
　　　　Visserman　searches the cause　itself with　the intertwinedness with
　　　　*linkse partijen'.*
　　　　left　　parties
　　　　'Vissermans searches the cause itself in the intertwinedness with the parties on the left.'
　　　　　　　(WR-P-P-G-0000150181.p.7.s.2)

　　b.　*Het Europese parlement zoekt　de verklaring　vooral in het*
　　　　The European parliament　searches the explanation mainly in　the
　　　　*democratisch tekort (...)*
　　　　democratic　deficit
　　　　'The European parliament searches the explanation mainly in the democratic deficit...'
　　　　　　　(WR-P-P-G-0000141214.p.2.s.7)

　　c.　*De Wilde zoekt　naar de oorzaken.*
　　　　De Wilde searches to　the causes
　　　　　　　(WR-P-P-G-0000715408.p.6.s.1)

　　d.　*Toen wetenschappers naar een verklaring zochten... ontdekten*
　　　　When scientists　　　to　an　explanation searched　discovered
　　　　*ze　cruciale gegevens over　de geschiedenis van Congo.*
　　　　they crucial data　　about the history　　of　Congo
　　　　When scientists were searching for an explanation,... they discovered crucial facts about the history of Congo.'
　　　　　　　(WR-P-E-G-0000004195.p.350.s.1)

As for *manier zoeken* 'search way', the hypothesis-generating analyses did not point to a relevant distinction, and the distinction based on LOCATIVE ADJUNCT also failed to be confirmed – albeit possibly due to the complete absence of instances with a locative adjunct in the hypothesis-testing set. Perhaps our failure to find any kind of possibly relevant distinction here is not surprising, given the nature of most instances of *manier* 'way', as in (190). These instances do not have much in common in terms of meaning: *manier* is rather used as a placeholder to introduce a subordinate clause.

190. a. *We zochten een originele manier om de aandacht van*
we searched an original way to the attention of
*sollicitanten op ons ziekenhuis te richten.*
applicants on our hospital to point
'We were looking for an original way to point the attention of applicants to our hospital.'
(WR-P-P-G-0000179354.p.4.s.4)

b. *De NAVO zoekt een manier om Oekraïne en Georgië*
The NATO searches a way to the_Ukraine and Georgia
*dichter bij NAVO-lidmaatschap te brengen, zonder Rusland voor*
closer with NATO-membership to bring without Russia for
*het hoofd te stoten.*
the head to bump
'NATO is looking for a way to bring the Ukraine and Georgia closer to a NATO-membership without stepping on the toes of Russia.'
(WS-U-E-A-0000092612.p.1.s.1)

c. *Om zijn hobby te financieren, zoekt hij voortdurend naar*
to his hobby to finance searches he continuously to
*manieren om aan extra geld te komen.*
ways to on extra money to come
'To finance his hobby, he is continuously searching for ways to get extra money.'
(WR-P-P-G-0000018974.p.6.s.3)

d. *Ik zoek naar een innovatieve manier om dat tot stand te*
I search to an innovative way to that to stand to
*brengen.*
bring
'I search for an innovative way to bring that about.'
(WR-P-P-G-0000127409.p.25.s.5)

The distinction in terms of AGENT TYPE for *dader zoeken* 'search perpetrator' and *spoor zoeken* 'search track/leads' links up with the distinction that was found for *slachtoffer zoeken* 'search victim' in Subsection 6.3.3. For *vorm* 'shape' and *weg* 'road' – and *woord* 'word' – we have found distinctions that only appear relevant for these specific themes.

How to interpret these results for *zoeken* 'search'? One possibility is to say that the distinction between 'seek to make/acquire' and 'look for' is too vague and abstract to account for the concrete meaning distinctions that we have now found on the lower levels – much like Section 6.4 concluded for the distinction in terms of directionality at the level of the preposition. This might be a bit quick on the trigger, however: after all, the distinction can directly account for the effect of LOCATIVE ADJUNCT for the explanation-group, as well as the effect of AGENT TYPE for the victim group and the effect of MEANING WEG for *weg zoeken* 'search road'. Still, we currently see no way it might have correctly predicted the effect of WORD SPECIFICITY for *woord zoeken* 'search word', nor the effect of MEANING VORM for *vorm zoeken* 'search form'. Hence, we propose to retain the distinction between 'seek to make/acquire' and 'look for' at the level of the verb, but also require more concrete constructions for *naar woord zoeken* 'be lost for words' and *naar vorm zoeken* 'try to get into top condition'. The idiosyncratic distinctions of these verb-theme combinations should then be specified for them individually.

## 7.4    Conclusion

The hypothesis-generating procedure based on Memory-based Learning proposed in this chapter has pointed towards several low-level distinctions that have been confirmed in subsequent hypothesis-testing, especially when the C-indexes of the analyses are high. It can hence generally be considered a success.

The proposed procedure may also be useful even if one is solely interested in testing a priori hypotheses at a high level of abstraction: it can be used to execute a quick sweep at the lower levels to check whether there are any strong local effects at play that may distort the results of a higher-level analysis. This would involve running many MBL-analyses at the lower levels, but only checking those where there is a peak in C-index relative to the other analyses. For these analyses, the gain ratios can then be inspected to track down a local distinction that may be causing the peak.

The following chapter will now shift the focus from semantic and lectal distinctions to the effect of language complexity, by putting the complexity hypotheses to the test.

# 8    Testing the complexity hypotheses

After testing lectal and semantic hypotheses in the previous chapters, we now turn to the complexity hypotheses, which are repeated below. Large parts of the present chapter are based on Pijpops et al. (2018), although the analyses differ to some extent, since they are based on a more developed understanding of the alternation. The chapter is structured as follows. Section 8.1 explains which data are used to test the hypotheses and why. Section 8.2 then checks whether the reasoning underlying the complexity hypotheses still holds for the selected data. Next, Section 8.3 presents the analyses and Section 8.4 discusses the conclusions.

**Com1**:  The Complexity Principle is primarily caused by constraints in language **production**. Therefore, there should be a negative correlation between THEME COMPLEXITY and the likelihood of *naar* 'to' when the theme precedes the verbs, and a positive correlation when the verb precedes the theme.

**Com2**:  The Complexity Principle is primarily caused by constraints in the language **channel**. Therefore, there should be a negative correlation between THEME COMPLEXITY and the likelihood of *naar* 'to' when the theme precedes the verbs, and a positive correlation when the verb precedes the theme.

**Com3**:  The Complexity Principle is primarily caused by constraints in language **comprehension**. Therefore, there should be a strong positive correlation between THEME COMPLEXITY and the likelihood of *naar* 'to' when the theme precedes the verbs, and a weaker positive correlation when the verb precedes the theme.

## 8.1    Data

Having done our best to track down any meaning differences between the variants in the previous chapters, we are now in a good position to attempt to control for these differences (cf. the end of Subsection 4.3.2.2). Our best shot at controlling for semantic differences is by using data of the verbs *peilen* 'gauge' and *zoeken* 'search'. The data of the verb *verlangen* 'desire' is not used, since this verb exhibited a clear meaning difference between the variants with outspoken lexical biases. By contrast, we have not managed to confirm any meaning difference for Belgian *peilen* 'gauge', although there were some indications that a distinction between gauging construed as directly judging and gauging construed as asking might be at play. Since this verb hardly exhibited any variation in the Netherlands, only the Belgian data will be used, like in Chapter 6.

As for *zoeken* 'search', we seemed to succeed in confirming a semantic distinction at the level of the verb, albeit one whose predictions at the level of the object were not consistently borne out, and whose effects appear weaker than for *verlangen* 'desire', as may be inferred from comparing the C-indexes in Tables 14-15 and 17-18 in Section 6.2, and a visual inspection of the effect plots in Figure 9a-10a and 12a-13a in the same section. To control for this distinction, we will use the measure based on relative distributional proximity employed in Chapter 6, as well as random effects for the theme roots. Because of this, the instances of the pronominal themes and the collexemes again had to be excluded, as in Section 6.2.

Furthermore, Section 6.3 and Chapter 7 unveiled several local effects at the level of the object for *zoeken* 'search'. To eliminate their influence, the instances of the theme roots where semantic distinctions were confirmed are likewise excluded from the analyses. These include *slachtoffer* 'victim', *woord* 'word', *oorzaak* 'cause', *verklaring* 'explanation', *dader* 'perpetrator', *spoor* 'track', *vorm* 'shape' and *weg* 'road'. The instances of *oplossing* 'solution' were also excluded, even though we did not confirm a distinction for this theme root. The reason is that Figure 32n in Section 7.2 did point to a potentially strong distinction, and we want to err on the side of caution in this regard.

Finally, the data of the other verbs, such as *grijpen* 'grab', *bellen* 'phone' etc. were not used, because these verbs were not investigated in the same depth as *verlangen* 'desire', *peilen* 'gauge' and *zoeken* 'search', and hence, some semantic distinctions may still have gone undetected. We will then test the complexity hypotheses on three datasets: Belgian *peilen* 'gauge', Belgian *zoeken* 'search' and Netherlandic *zoeken* 'search'.

## 8.2 Checking the rationale behind the complexity hypotheses

The rationale behind all three complexity hypotheses is based on the argument that when the theme precedes the verb, the use of the preposition *naar* 'to' restricts the choice of verb. We first turn to Com1 and Com3. Because *naar* limits the choice of verb, Com1 predicted that producers will be more inclined not to use the preposition as the theme becomes more complex. Not using *naar* will enable them to postpone the choice of verb until after the completion of the theme. Meanwhile, Com3 argued that language comprehenders would benefit more from the presence of the preposition in front of complex themes when the theme precedes the verb, because the presence of the preposition would enable to easily infer the upcoming verb.

Objecting to this line of reasoning, it could be claimed that the addition of *naar* does not actually make the verb that much more predictable when the theme precedes the verb. After all, the theme by itself already narrows down the list of possible verbs to a large degree, and perhaps *naar* doesn't contribute much to narrowing it down even further. In that case, the rationale behind Com1 and Com3 would not hold up. We can then ask, for all instances in our datasets where the theme precedes the verb, how much more predictable *naar* would actually make the verb if it were included. We estimate this in the following way.

For each theme root in the subset of the Belgian *peilen*-data where the theme precedes the verb, we count the number of times it appears as the syntactic head of a noun phrase with *peilen* 'gauge' in the Belgian part of the Sonar-corpus. Next, we count the number of times it appears as the syntactic head of a noun phrase with any verb in the same subcorpus. Finally, we divide the former by the latter. This yields for each theme root the probability that it combines with *peilen* 'gauge' in Belgium. The average of these probabilities over data where the theme precedes the verb is 0.0033. This means that, once we know the theme root, there's on average a probability of 0.33% that the upcoming verb is *peilen* 'gauge'.

We now do the same calculations for the prepositional variant. We count for each theme root in the same subset, the number of times it appears as the syntactic head of a prepositional phrase headed by *naar* 'to' with *peilen* 'gauge', in the Belgian part of the Sonar-corpus. Next, we count the number of times it appears as the syntactic head of a prepositional phrase headed by *naar* 'to' with any verb. Finally, we divide the former by the latter. The average of these probabilities over the same subset is 0.2401. This means that, given the theme root and the preposition naar, there's on average a 24.01% probability that the verb will be *peilen* 'gauge'. Including *naar* 'to' can thus make comprehenders on average 73.76 times more

confident in their guess that the following verb is *peilen* 'gauge', which we regard as a considerable increase.[79]

When we do the same calculations for *zoeken* 'search', we find that, given the theme root, there's on average a 2.80% probability that the upcoming verb is *zoeken* 'search' in the Belgian data, while given the theme root and the preposition *naar* 'to', there is on average a probability of 21.43% chance that *zoeken* 'search' follows. As for the Netherlandic data, we find that there is an average probability of 3.64% that the verb is *zoeken* 'search' given the theme root, versus an average probability of 20.98% given the theme root and the preposition. In sum, the addition of *naar* 'to' consistently makes the verb considerably more predictable. The rationale behind Com1 and Com3 thus holds up.

We now turn to the channel-driven hypothesis Com2, which is also fundamentally based on the same assumption that when the theme precedes the verb, the preposition limits the choice of verb. Although the following calculations may appear to take a different outlook on the matter, it will be shown that they are fundamentally the same as above.

These calculations are based on Jaeger (2010: 28), who, when studying the English *that*-alternation as in (191), estimates the Shannon information of a complement clause in a similar way. He reasons that the probability of a following complement clause – and hence its information content – is dependent on the choice of matrix verb. That is, the probability of a following complement clause is very high for some matrix verbs, such as *think*, and lower for other, such as *confirm*. Hence, the information content of the complement clause is lower when the matrix verb is *think* than when it is *confirm*. He then estimates this information content by taking the negative base 2 logarithm of the probability that a complement cause would follow given the matrix verb root. These are necessarily only approximate measurements (Jaeger 2010: 28).

191. a.  *My boss confirmed we were absolutely crazy.*
   b.  *My boss confirmed that we were absolutely crazy.*
             (Taken over from Jaeger 2010: 27)

The channel-driven hypothesis Com2 argued that the preposition *naar* 'to' is informationally heavier when the theme precedes the verb than when the verb precedes theme. In other words, $\Delta I_{\text{naar}}$ in Equation 1 should be positive. We can estimate this difference in the following way.

Say that all sentences in (192) express the same proposition, i.e. that they all contain the same information in total, albeit spread out differently. Now we want to estimate the difference in Shannon information of *naar* when the theme precedes the verb, as in (192a), versus when the verb precedes the theme as in (192b), as

---

[79] The increase with factor 73.76 is calculated on the unrounded probabilities.

Equation 1 expresses in (i). This difference corresponds to the degree in which *naar* 'to' in (192a) makes the verb more predictable.[80] That is, it corresponds to the difference in information of *zoeken* in (192a) versus in (192c), as expressed in (ii).

We now assume that it is primarily the theme root that makes *zoeken* in instances like (192c) more predictable and we therefore estimate the information of *zoeken* in (192c) as its information given the theme root. Correspondingly, we estimate the information of *zoeken* in (192a) as its information given the theme root and *naar*. We now have (iii). Shannon Information can be calculated as the negative logarithm of the probability, which gives us (iv).[81] The formula in (iv) now contains the probabilities as they were calculated above. We fill these in, which yields the estimated $\Delta I_{naar}$. Finally, we take the average of these values over the dataset at issue. This yields the averages in (v).[82]

(i)     $\Delta I_{naar}$    $= I(naar \mid \textit{preverbal theme}) - I(naar \mid \textit{postverbal theme})$

(ii)           $= I(zoeken \mid \textit{preverbal nominal theme}) -$
            $I(zoeken \mid \textit{preverbal prepositional theme})$

(iii)         $\approx I(zoeken \mid \textit{preverbal theme root}) -$
            $I(zoeken \mid \textit{preverbal theme root} \ \& \ naar)$

(iv)         $= -log_2 \, p(zoeken \mid \textit{preverbal th.root}) +$
            $log_2 \, p(zoeken \mid \textit{preverbal th.root} \ \& \ naar)$

(v)      $\textit{Average} \ \Delta I_{naar} \ \textit{for Belgian}$ peilen $\textit{'gauge'} \approx 7.36 \ \textit{bits}$
        $\textit{Average} \ \Delta I_{naar} \ \textit{for Belgian}$ zoeken $\textit{'search'} \approx 3.65 \ \textit{bits}$
        $\textit{Average} \ \Delta I_{naar} \ \textit{for Netherlandic}$ zoeken $\textit{'search'} \approx 3.82 \ \textit{bits}$

Equation 1: Estimation of the difference in information content of the preposition when the theme precedes the verb versus when the verb precedes the theme.

---

[80] Of course preverbal *naar* in (192a) also makes the theme root more predictable, but this holds equally for postverbal *naar* in (192b).

[81] We used a logarithm with base 2, as in the seminal paper (Shannon 1948).

[82] The results in (v) of Equation 1 are not equal to the difference between the negative logarithms of the average probabilities mentioned above. The reason is technical: the calculations in Equation 1 first take the negative logarithm of the probabilities and then average over them, rather than the other way round. Taking the average of logarithms is not equal to taking the logarithm of averages. Neither strategy is necessarily better than the other, however: it is best to do whatever makes most sense from a theoretical point of view for the case study at hand. Taking the channel perspective, what really matters is not the probabilities as such, but rather the information density in the sentence. In other words, the information is more 'real' than the probabilities. As such, Equation 1 takes the average of the information differences, rather calculating the information difference based on the average probabilities.

192. a. *Naar evenwicht zochten    mijn armen.*
        to   balance    searched   my   arms

     b. *Mijn armen zochten    naar evenwicht.*
        my   arms   searched   to    balance

     c. *Evenwicht  zochten    mijn armen.*
        balance        searched   my   arms
        'My arms were searching for balance.'
                (WR-P-P-G-0000126206.p.81.s.1)

The positive values in (v) indicate that the information content of the preposition *naar* 'to' is on average several bits higher when the theme precedes the verb than when the verb precedes the theme. If the verb precedes the theme, the preposition evidently cannot contain any information about the verb, since the verb is already known when *naar* 'to' is heard or read. Since *naar* 'to' is thus informationally light, it can nicely combine with complex, informationally heavy themes. Meanwhile, if *naar* 'to' precedes the verb, it is burdened with a large chunk of the information content otherwise contributed by the verb, thus rendering it informationally heavy. In that case, it would be preferable not to combine it with a complex, informationally heavy theme. The rationale behind the prediction Com2 thus holds up.

## 8.3    Analyses

We can now proceed with the analyses to test the complexity hypotheses. In these analyses, we want to control for semantic differences as mentioned above, but also for two correlates of VERB-THEME ORDER, viz. CLAUSE TYPE and VERB FINITENESS. CLAUSE TYPE is a categorical variable that distinguishes between main clauses and subordinate clauses; VERB FINITENESS is also categorical and discerns between instances where the main verb, i.e. *peilen* 'gauge' or *zoeken* 'gauge', is a finite form, an infinitive or a participle.[83] We do not have any hypotheses underlying these variables – apart from the complexity hypotheses themselves, which are more straightforwardly operationalized by VERB-THEME ORDER – and hence did not include them in any previous models. However, since an effect of VERB-THEME ORDER could easily be caused by an underlying effect of CLAUSE TYPE or VERB

---

[83] Instances for which the values of these variables could not be drawn from the Alpino-parses were manually annotated for them.

FINITENESS (cf. Table 12 in Subsection 4.3.2.3), we do want to control for them in the present analyses.

VERB FINITENESS is a categorical variable whose levels contain an internal structure. That is, both infinitives and participles are non-finite forms, and neither can be placed in the first verbal pole, unlike the finite forms. For implementing such structured categorical variables, the use of dummy coding or treatment contrasts is not ideal. The reason is that in this implementation, one of the levels needs to be arbitrarily chosen to act as a reference level, to which the other levels are individually contrasted. For instance, if the level finite is chosen as a reference level, then the finite forms would be on the one hand contrasted with the infinitives, and on the other hand with the participles, but the finite forms would not be contrasted with the non-finite forms, nor would the infinitives and the participles be contrasted with each other.

Therefore, we instead opt for user-defined sum-to-zero contrasts, also called user-defined sum coding. This allows us to contrast on the one hand finite forms with non-finite forms, and on the other hand, participles with infinitives among the non-finite forms. In practice, it means that VERB FINITENESS is implemented through two parameters. The first is set to 1 for finite forms, and -0.5 for infinitives and participles. The second parameter is set to 0 for finite forms, to 1 for infinitives, and -1 for participles. The variables VERB-THEME ORDER and CLAUSE TYPE are then also implemented through sum-to-zero contrasts, although they each only require a single parameter: for *theme-verb*, the respective parameter is set to -1, for *verb-theme* to 1; the parameter of CLAUSE TYPE is set to -1 for main clauses, and 1 for subordinate clauses.

We begin the analyses with Belgian *peilen* 'gauge'. For this verb, we want to build a regression model with fixed effects for SEMANTIC COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION, THEME COMPLEXITY, VERB-THEME ORDER, CLAUSE TYPE, VERB FINITENESS, and an interaction between THEME COMPLEXITY and VERB-THEME ORDER, as well as random effects for THEME ROOT and CORPUS COMPONENT.[84] These fixed effects require 7 parameters in total. However, once we exclude the collexemes of SEMANTIC COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION (cf. Equation 3 in Subsection 4.2.2.2.2), we are only left with sufficient transitive occurrences for 5 parameters, following the rule of thumb not to include more parameters in a regression model than the number of data points of the least frequent response level divided by 20.

It was therefore decided to drop SEMANTIC COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION, since (i) it yielded a non-significant effect in Table 16 in Subsection 6.2.2 anyway, and (ii) the random effect THEME ROOT can still partially control for

---

[84] The hypothesized effect of OBJECTAL COHERENCE TO THE *NAAR*-CONSTRUCTION was not confirmed in any of the regression models in Chapter 6, so it is kept out of the present analyses.

a meaning difference.[85] Retaining CLAUSE TYPE and VERB FINITENESS was hence considered more crucial than retaining SEMANTIC COHERENCE TO THE *PEIL-NAAR*-CONSTRUCTION. Lastly, removing SEMANTIC COHERENCE TO THE PEIL-NAAR-CONSTRUCTION comes with the additional benefit that the instances of the collexemes no longer need to be excluded from the dataset, nor the instances with theme roots for which no vector could be calculated. The reasons for their exclusion were respectively that they would render SEMANTIC COHERENCE TO THE PEIL-NAAR-CONSTRUCTION circular, and that the variable had no value for them. These reasons no longer apply when SEMANTIC COHERENCE TO THE PEIL-NAAR-CONSTRUCTION is not taken up in the model. As such, the model can be based upon more data. All theme roots that occurred only once in the dataset were again binned into a rest category for the random effect THEME ROOT, as in Chapter 6.

This model was then fitted to the data. The variance inflation factor of VERB-THEME ORDER in this model was 6.13, which is higher than the conservative threshold of 5 – although lower than the more permissive threshold of 10 (Levshina 2015: 160).[86] VERB FINITENESS and CLAUSE TYPE also yield high variance inflation factors of 3.87 and 3.45, respectively. Therefore, VERB FINITENESS was removed from the model, which caused the variance inflation factors to drop below 5. The specifications of this new model can be found in Table 21, while Figure 34 shows the effect plot of the interaction between THEME COMPLEXITY and VERB-THEME ORDER.

The Appendix reports the specifications of the original model, containing both VERB FINITENESS and CLAUSE TYPE, in Table 28, as well as its effect plot showing the interaction between THEME COMPLEXITY and VERB-THEME ORDER, in Figure 38. In addition, Table 29 and Figure 39 in the Appendix show the specifications and effect plot of a model from which CLAUSE TYPE is removed, but VERB FINITENESS is retained, thereby also bringing the variance inflation factors below 5. The results of these models are not qualitatively different from those in Table 21 and Figure 34.

---

[85] We say *partially*, because the lexical biases of the theme roots would actually not be random if they were indeed caused by a meaning difference. Still, this is the best we can do, given the convergence issues.

[86] All variance inflation factors are calculated on models where the interaction was removed.

AIC: 636.3          Transitive observations: 163
C-index: 0.863      Prepositional observations (success level): 847

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | 1.98 | 0.42 | 4.70 | < 0.0001 |
| THEME COMPLEXITY | | 0.00 | 0.17 | 0.00 | 0.9966 |
| VERB-THEME ORDER | *theme-verb* (vs. *verb-theme*) | -0.63 | 0.30 | -2.13 | 0.0329 |
| CLAUSE TYPE | *main clause* (vs. *sub.clause*) | 0.60 | 0.19 | 3.17 | 0.0016 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *theme-verb* (vs. *verb-theme*) | -0.25 | 0.16 | -1.55 | 0.1211 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| THEME ROOT | 143 | 2.64 | 1.63 |
| CORPUS COMPONENT | 11 | 0.09 | 0.30 |

Table 21: Specifications of a regression model fitted on the Belgian instances of the verb *peilen* 'gauge' to test the complexity hypotheses.



a. *theme-verb*          b. *verb-theme*
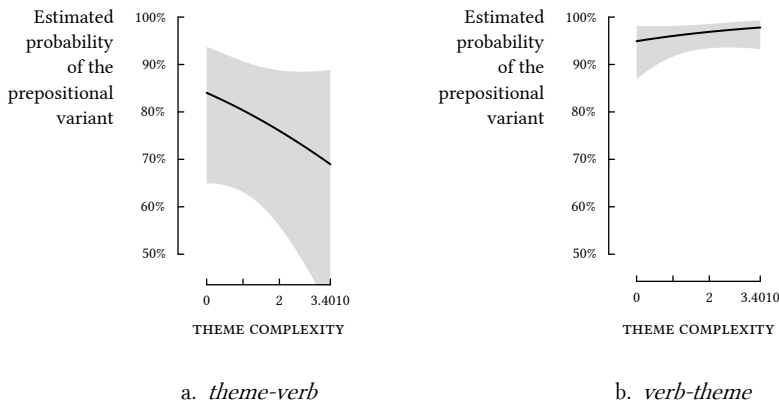
Figure 34: Effect plot of the interaction between THEME COMPLEXITY and VERB-THEME ORDER in the regression model of Belgian *peilen* 'gauge' in Table 21.

We now turn to Belgian *zoeken* 'search'. Here, we build a regression model with the fixed effects SEMANTIC COHERENCE TO THE *ZOEK-NAAR*-CONSTRUCTION, THEME COMPLEXITY, VERB-THEME ORDER, CLAUSE TYPE, VERB FINITENESS, an interaction between THEME COMPLEXITY and VERB-THEME ORDER, and well as random effects for THEME ROOT and CORPUS COMPONENT. The theme roots that only occurred once or

twice in the dataset were again binned into a rest category, and a more coarse-grained classification of the corpus components was used, as in Chapter 6. This model was then fitted to the data and its specifications can be found in Table 22, while the effect plot for the interaction is shown in Figure 35. The variance inflation factors are again high, viz. 4.43, 2.54 and 2.73 for respectively VERB-THEME ORDER, CLAUSE TYPE and VERB FINITENESS, but at least rank under the threshold of 5.

For Netherlandic *zoeken* 'search', the same variables were selected and the specifications of this model are listed in Table 23. Figure 36 shows the effect plot of the interaction between THEME COMPLEXITY and VERB-THEME ORDER. The variance inflation factors are also high, with 4.02, 2.69 and 2.29 for respectively VERB-THEME ORDER, CLAUSE TYPE and VERB FINITENESS, but under the threshold of 5.

AIC: 24,557.6          Transitive observations: 28,843
C-index: 0.809         Prepositional observations (success level): 4,918

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | -2.25 | 0.21 | -10.80 | < 0.0001 |
| SEMANTIC COHERENCE TO THE *ZOEK-NAAR*-CONSTRUCTION | | 1.53 | 0.14 | 11.12 | < 0.0001 |
| THEME COMPLEXITY | | -0.07 | 0.03 | -2.77 | 0.0056 |
| VERB-THEME ORDER | *theme-verb* (vs. *verb-theme*) | -0.01 | 0.04 | -0.24 | 0.8091 |
| CLAUSE TYPE | *main clause (vs. sub. clause)* | 0.08 | 0.03 | 2.55 | 0.0109 |
| VERB FINITENESS | *finite* (vs. *non-finite*) | 0.04 | 0.04 | 1.06 | 0.2896 |
| | *infinitive (vs. participle)* | -0.46 | 0.04 | -12.44 | < 0.0001 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *theme-verb* (vs. verb-theme) | -0.25 | 0.03 | -9.80 | < 0.0001 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| THEME ROOT | 1538 | 1.43 | 1.20 |
| CORPUS CATEGORY | 4 | 0.14 | 0.37 |

Table 22: Specifications of a regression model fitted on the Belgian instances of the verb *zoeken* 'search' to test the complexity hypotheses.

a.  *theme-verb*     b.  *verb-theme*
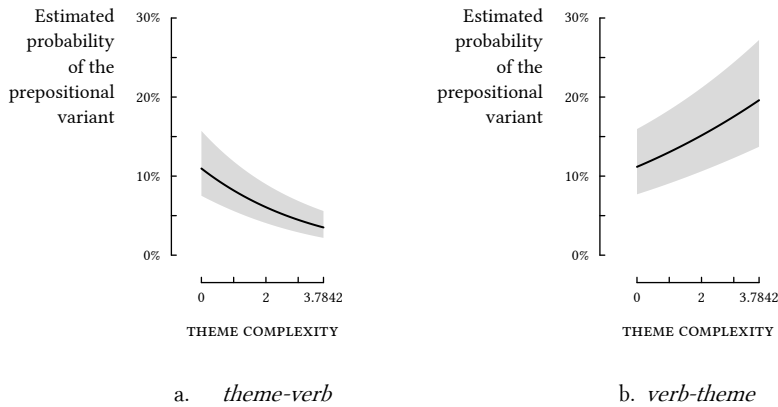
Figure 35: Effect plot of the interaction between THEME COMPLEXITY and VERB-THEME ORDER in the regression model of Belgian zoeken 'search' in Table 22.

AIC: 9,921.1          Transitive observations: 9,990
C-index: 0.829        Prepositional observations (success level): 2,233

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | Intercept | -2.02 | 0.16 | -12.61 | < 0.0001 |
| SEMANTIC COHERENCE TO THE *ZOEK-NAAR-* CONSTRUCTION | | 1.39 | 0.15 | 9.33 | < 0.0001 |
| THEME COMPLEXITY | | -0.07 | 0.04 | -1.94 | 0.0519 |
| VERB-THEME ORDER | *theme-verb* (vs. *verb-theme*) | 0.14 | 0.06 | 2.12 | 0.0345 |
| CLAUSE TYPE | *main clause* (vs. *sub. clause*) | 0.33 | 0.05 | 7.26 | < 0.0001 |
| VERB FINITENESS | *finite* (vs. *non-finite*) | 0.22 | 0.06 | 3.75 | 0.0002 |
| | *infinitive* (vs. *participle*) | -0.26 | 0.06 | -4.44 | < 0.0001 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *theme-verb* (vs. verb-theme) | -0.26 | 0.04 | -7.12 | < 0.0001 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| THEME ROOT | 732 | 1.47 | 1.21 |
| CORPUS CATEGORY | 4 | 0.05 | 0.22 |

Table 23: Specifications of a regression model fitted on the Netherlandic instances of the verb *zoeken* 'search' to test the complexity hypotheses.
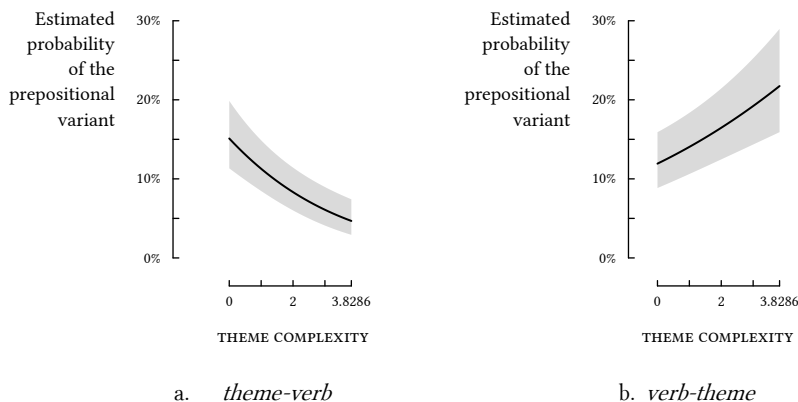
Figure 36: Effect plot of the interaction between THEME COMPLEXITY and VERB-THEME ORDER in the regression model of Netherlandic *zoeken* 'search' in Table 23.

All effect plots in Figure 34-36 show a negative correlation when the theme precedes the verb and a positive correlation when the verb precedes the theme. This confirms the production-driven and channel-driven complexity hypotheses.

## 8.4    Conclusions

We have found that, at least with regard to the Dutch transitive-prepositional alternation, the correlation between complexity and explicitness described in the Complexity Principle appears to be primarily motivated by either production processing or channel constraints. This dovetails with the majority of the literature on the influence of production vs. comprehension processing, including Ferreira and Dell (2000), Kraljic and Brennan (2005), Roland, Elman and Ferreira (2006), Levy and Jaeger (2007), Gennari and Macdonald (2009), MacDonald and Thornton (2009), Jaeger (2010), Ferreira and Hudson (2011), Ferreira and Schotter (2013) and MacDonald (2013).

Still, it should be noted that our results do not entail that the use of explicit coding is completely unbeneficial to the comprehender. Regarding the production perspective, we included both the PDC-model and the collateral signals account, both of which hold that the comprehender does benefit, albeit in an indirect way. There is, in fact, strong evidence that the comprehender interprets disfluencies and grammatical markers such as the English subordinator *that* and Dutch existential *er* as signals of upcoming production difficulties or unpredictable material (Jaeger 2005; Grondelaers et al. 2009; Clark and Fox Tree 2002; Corley and Hartsuiker

2003; Collard et al. 2008). Perhaps it is more relevant for the comprehender to receive notifications on the current state of production than to procure sentences that are easier to parse. Additional research would need to confirm whether the comprehender does indeed interpret the preposition *naar* 'to' as such a signal. Our current results certainly do not exclude this; instead, they show, in fact, that *naar* 'to' can function as a reliable signal of production difficulties or unpredictable material.

Regarding the channel perspective, it may very well be in the interest of the comprehender to burden his or her own cognitive processing, if a more important goal is safeguarded. For example, it is both in the interest of the comprehender and the producer to make sure that as little information as possible is lost in the noisy language channel by making sure the information density does not exceed its optimal level too much or too often (Fenk and Fenk-Oczlon 1993; Jaeger 2010; Collins 2014). If that leads to tendencies that require more cognitive effort during parsing, this may very well be a price worth paying.

Meanwhile, the findings also indicate that the Complexity Principle should not be interpreted as a blind law, but rather as a general tendency that holds in most, but not all contexts. This is also argued by Rohdenburg (2016) and Willems and De Sutter (2015), who propose further refinements to the Complexity Principle. In order to determine in which context we can expect the principle to hold, we need to consider its underlying mechanism, as well as the specifics of the case study. For example, we have shown that the order of theme and verb is a relevant distinction in our case study, with the effect of Complexity Principle reversing when the theme precedes the verb. Such context-determined restrictions to the Principle present a possible caveat for alternation studies, which do not always take the underlying mechanisms of the Complexity Principle into account.

With the completion of the present Chapter, the analyses of the data in this thesis are concluded. Chapter 9 will end this thesis by formulating answers to the general research questions and returning to some of the points raised in Chapter 2.

# 9     Conclusions and discussion

Now that all the results of this thesis have been presented, the present chapter will review how they fit together in the smaller and bigger puzzles we have set out to tackle, and what general conclusions can be drawn. First and foremost, Sections 9.1 and 9.2 set out to provide an answer to the three questions that sparked this investigation, viz. how, why and where does argument structure vary with respect to the Dutch transitive-prepositional alternation. Based on this, Section 9.3 returns to the various definitions of an alternation presented in Chapter 2, and discusses which segments of our case study best fit each definition and to what degree. Finally, Section 9.4 suggests a number of promising avenues for future research and Section 9.5 concludes this thesis by summarizing its main contributions to various fields of study.

## 9.1     How and why does argument structure vary?

This section reviews the descriptive findings of this thesis in the order that they have been presented, starting with the individual verbs *verlangen* 'desire', *peilen* 'gauge' and *zoeken* 'search', and then turning to telephonic verbs and finally to the motoric and venatic verbs. In doing so, it discusses the answers to the how and why-questions.

We start with the how-question for the verb *verlangen* 'desire'. Here, we have found a very clear semantic difference between both variants: when the desire in question is construed as a demand, the transitive variant is chosen, while when the desire is construed as a longing, language users employ the prepositional variant. This results in a nicely balanced distribution between the variants at the level of the verb, while creating outspoken lexical biases at the level of the object. Within this semantic distinction, a lectal effect also seems at play: the semantic distinction appears more strict in the Netherlands than in Belgium.

Why is this the case? As for the difference between Belgium and the Netherlands, its origin is plausibly found in a higher degree of functional specialization of linguistic variation in the Netherlands, due to the delayed

standardization of Dutch in Belgium (cf. Section 4.1). In other words, the centuries of enduring, gradual standardization in the Netherlands would have caused the available linguistic variation to become locked down into lexical biases or clear semantic distinctions (Grondelaers, Speelman and Geeraerts 2008). Meanwhile, this would be less the case in Belgian Dutch, where the standardization process only really took off in the 20th century.

Meanwhile, the semantic distinction may have developed through the lexical origin mechanism operating at the verb-level. This cannot be the whole story however, since the lexical origin mechanism itself still needs to latch on to some initial lexical biases. This then pushes the why-question further: what may have given rise to the initial lexical biases that sparked the lexical origin mechanism?

One possibility has already been proposed in Figure 4 in subsection 4.2.2.2.1, viz. the influence of complexity. That is, perhaps themes that imply a 'longing for' are more often used in complex noun phrases, which would lead to ostensible lexical biases for such themes. Another possibility is an effect of constructional contamination from the nominalization of the verb (Pijpops and Van de Velde 2016; Hilpert and Flach forthc.). On the one hand, there is reason to believe that the nominalization *een verlangen* 'a desire' would have a preference for the prepositional variant: in the transitive variant, the theme would need to be realized in a prepositional phrase headed by the preposition *van* 'of', as in (193). Since the agent can also be realized in a *van*-phrase, as in (194a-c), this in principle creates a potential ambiguity and makes it hard to express both agent and theme for the same nominalization. In the prepositional variant, the theme is realized in a *naar*-phrase, as in (194b-d), as a result of which these problems are avoided.[87]

On the other hand, a quick corpus survey shows that the nominalization typically expresses a longing. As such, things that are often longed for, such as life, death, rest, etc. would more often appear in the nominalization than things that are demanded. This would in turn cause language users to often hear *het verlangen naar een leven* 'the desire for a life', *het verlangen naar rust* 'the desire for rest', etc., while hearing less often *het verlangen naar tegenprestaties* 'the desire for counter effors' or *het verlangen naar offers* 'the desire for sacrifices'. The nominalizations show a superficial resemblance to prepositional instances of the verb *verlangen* 'desire', which could trigger the mechanism of constructional contamination and cause the same lexical biases to develop among the verbal instances.

Concretely, theme roots like *leven* 'life' and *rust* 'rest' would come to prefer the prepositional variant among instances of the verb *verlangen* 'desire', because

---

[87] Granted, a consistent choice for the preposition *door* 'by' to express the subject rather than *van* 'of' would also avoid these problems, as in *het verlangen door Janek van het zigeunermeisje Zefka*, lit. 'the desire by Janek of the gipsy girl Zefka' or *het verlangen van mensenoffers door de goden*, lit. 'the desire of human sacrifice by the gods'.

language users so often hear the string *verlangen naar een leven* 'desire to a life' etc. in the nominalization. Meanwhile, the same effect would not emerge for theme roots like *tegenprestatie* 'counter effort' and *offer* 'sacrifice', because these would only rarely appear in the nominalization.[88]

This then pushes the why-question even further: why is the nominalization *een verlangen* 'a desire' more often used to express a longing than a demand? Nominal alternatives exist both to express a demand, e.g. *een eis* 'a demand, a claim' and to express a longing, e.g. *een hunkering* 'a longing, a yearning'. Perhaps *hunkering* 'a longing, a yearning' is dispreferred for some reason; we do not know an answer to this final question yet. Of course, all of this is mere speculation for the moment; a diachronic study is required to get to the bottom of it (cf. Geeraerts 1988: 660–661).

193. a.   *Er komt een officieel onderzoek naar het doden van de*
there comes a official investigation to the killing of the
*Hooglanders.*
Highlanders
'There will be an official investigation into the killings of the Highlanders.'
      (WS-U-E-A-0000022568.p.19.s.1)

    b.   *? Het verlangen van acceptatie zie je bij alle kinderen.*
the desire of acceptation see you with all children
'The desire for acceptation can be seen among all children.'

194. a.   *Er is een enorme spanning tussen het verlangen van de*
there is an enormous tension between the desire of the
*kinderen en de kansen die de maatschappij hen biedt.*
children and the chances that the society them offers
'There is an enormous tension between the desire of the children and the chances that society offers them.'

---

[88] In this sense, constructional contamination could in principle explain the positive correlation between SEMANTIC COHERENCE TO THE *VERLANG-NAAR*-CONSTRUCTION and the probability of the prepositional variant, even in absence of any semantic differentiation between the variants. It cannot, however, explain differences at the object level, as presented in Subsection 6.3.1.

b. *Er is een enorme spanning tussen het verlangen van de*
there is an enormous tension between the desire of the
*kinderen naar acceptatie en de kansen die de maatschappij*
children for acceptation and the chances that the society
*hen biedt.*
them offers
'There is an enormous tension between the desire of the children for acceptation and the chances that society offers them.'
(WR-P-P-G-0000015184.p.7.s.3)

c. *Het verlangen van Janek naar het zigeunermeisje Zefka blijft*
the desire of Janek to the gipsy_girl Zefka stays
*romantisch abstract.*
romantically abstract
'The desire of Janeck for the gipsy girl Zefka remains romantically abstract.'
(WR-P-P-G-0000057776.p.8.s.2)

d. *Het verlangen naar seks was opgelaaid.*
the desire for sex was flared_up
'The desire for sex had flared up.'
(WR-P-P-B-0000000208.p.930.s.1)

For the verb *peilen* 'gauge', we found a commanding lectal distinction: the prepositional variant is nearly non-existent in the Netherlands, while it is by far the dominant variant in Belgium. Within Belgium, both a semantic and processing effect appear to be at play, though we lack the data to conclusively confirm either. The lectal distinction fits our lectal hypothesis at the level of the preposition: individual verbs were expected to have stronger lexical biases in the Netherlands than in Belgium (cf. Subsection 4.1.2). The processing effect confirmed our production-based and channel-based hypotheses.

As for the semantic effect, the transitive variant seems to incline towards 'directly judge', while the prepositional variant tends more towards 'ask about'. To explain this, we might point towards the influence of *vragen naar* 'ask about', which would be synonymous with prepositional *peilen naar* 'ask about'. *Vragen naar* 'ask about' might have affected *peilen* 'gauge' either through direct analogy, or again through a combination of constructional contamination and the lexical origin mechanism. That is, because of *vragen naar* 'ask about/after', theme roots that are often asked about, e.g. expectations and satisfaction, would often appear behind the preposition *naar* 'to', which could also cause them to prefer the prepositional variant among instances of *peilen* 'gauge'. Once such lexical biases

are in place, the lexical origin mechanism could then take over to induce a semantic distinction.

Turning to the verb *zoeken* 'search', our analyses revealed a processing effect and what seemed to be various semantic distinctions. The processing effect confirmed the production and channel-based hypotheses, both in Belgium and the Netherlands. As for semantics, the analysis in Chapter 6 at the level of the verb indicated that objects that came into being by the act of searching preferred the transitive variant, whereas concrete objects that were literally looked for preferred the prepositional variant.

In addition, the analyses at the level of the object in Chapters 6 and 7 revealed more specific distinctions for particular verb-theme combinations. For the explanation-group, notably *verklaring zoeken* 'search explanation' and *oorzaak zoeken* 'search cause', as well as *waarheid zoeken* 'search truth' and perhaps *oplossing zoeken* 'search solution', the presence of a locative adjunct was shown to be relevant (cf. (200), (201) and (206) in Chapter 7). For the victim-group, i.e. *slachtoffer zoeken* 'search victim', *dader zoeken* 'search perpatrator' and *spoor zoeken* 'search lead', the crucial distinction seems to lay in the agent participant. If the search was performed by the authorities, the prepositional variant was triggered, while other forms of searching were expressed in the transitive variant (cf. (195) in Chapter 6).

Furthermore, for *vorm zoeken* 'search forms', instances of sportspeople trying to get into condition were expressed in the prepositional variant, while other instances preferred the transitive variant; for *woord zoeken* 'search words', the prepositional variant more often involved people trying to formulate a proposition, while literally looking for words was more often expressed in the transitive variant; for *weg zoeken* 'search road', the proposition of people trying to find their proper place is typically expressed in the transitive variant, while other forms of searching a road or a way are preferably expressed in the prepositional variant.

It may be argued that these low level distinctions are merely fixed collocations that can or even should be ignored in the study of argument realization.[89] However, this would mean that researchers would be a priori blind to any influence of the lower levels of abstraction on argument realization. Moreover, especially in a usage-based approach, such semi-fixed collocations should be taken seriously in any case, since they constitute a substantial part of language usage (Altenberg 1998; Erman and Warren 2000; Dąbrowska 2006, 2014: 625–627). Finally, we find such low-level distinctions both with the transitive variant, e.g. *weg zoeken* 'find one's place', and the prepositional variant, e.g. *naar vorm zoeken* 'try to get into top condition'. Even if we accept that these low-level distinctions are 'merely' fixed collocations, we can still reasonably ask why such a semi-fixed collocation

---

[89] Of course, this claim is only possible in approaches that employ a strict separation of lexicon and grammar (Geeraerts 2010a: 76, 82).

developed in the transitive variant for *weg* 'road' and in the prepositional variant for *vorm* 'form'.

Put more generally, how can we explain that the mere presence or absence of the same preposition seems to express so many different things, even among instances of a single verb, instead of just instantiating a single semantic distinction? Here follows a tentative, post-hoc explanation. The influence of complexity makes that only complex constituents or constituents carrying a lot of information should be preceded by a preposition. This influence could cause the distinction among the explanation-group, as argued in Section 7.3. Among the themes of this group, the presence of a locative adjunct was shown to be correlated with the use of the prepositional variant. It could possibly also lead to the distinction for *woord zoeken* 'search word': instances where the agent is trying to formulate a proposition, as in (195a), rarely contain other pieces of information; by contrast, instances where a word is literally sought, e.g. (195b) may specify exactly which word should be sought, how it should be sought, etc. The same holds for *vorm zoeken* 'search form': in sentences where sportspeople are trying to get into condition, the focus is typically on the *vorm* 'form/condition', as in (196a). In other sentences, such as (196b), the theme root *vorm* 'form' often functions as a mere placeholder, while the relevant information is somewhere else in the sentence.

195. a. *Hij zoekt naar de juiste woorden, (...) maar komt niet verder*
    he searches to the right words but comes not further
    *dan de uitlating dat het geen 'fair play-situatie' is geweest.*
    than the utterance that it no fair play situation is been
    'He searches for the right words, (...) but gets no further than the
    utterance that it hadn't been a "fair play situation".'
        (WR-P-P-G-0000030653.p.2.s.3)

   b. *Ze zoekt dan ook een heel moeilijk woord: gedwongen*
    she searches then also a very difficult word forced
    *huwelijk.*
    marriage
    'For she is looking for a very difficult word: forced marriage.'
        (WR-P-P-G-0000663990.p.2.s.3)

196. a. *Hellebaut zoekt naar de vorm voor het WK indoor en speelt*
    Hellebaut searches to the form for the WK indoor and plays
    *opnieuw met de lat op 2 meter hoogte.*
    again with the lat on 2 meter height
    'Hellebaut is trying to find the proper condition for the WK indoor and
    is again playing with the bar on a height of 2 meters.'
        (WS-U-E-A-0000097726.p.1.s.4)

b. *Om te overleven, zoeken verscheidene producenten nieuwe*
To to survive search various producers new
*vormen van samenwerking met concurrenten.*
forms of cooperation with competitors
'To survive, various producers are searching for new forms of
cooperation with competitors.'
(WR-P-P-H-0000047725.p.2.s.3)

This may in turn lead to a tendency towards quasi-noun-incorporation in the transitive variant (Booij 2009). We see this in the strong lexical biases towards the transitive construction for a.o. the theme roots *contact* 'contact', *steun* 'support' and *beschutting* 'cover' as in (197)-(199) (cf. Section 5.1). These theme roots do occur in the prepositional variant, as in the b-sentences, but much more rarely than in the transitive variant. Finally, such a tendency would lead to semantic bleaching of transitive *zoeken* from 'search' to 'seek to make/acquire', while prepositional *zoeken naar* would retain its compositional meaning 'look for'. This could in turn cause the distinctions for the victim-group and for *weg zoeken* 'search road'. Again, this is merely a post-hoc explanation and we would need diachronic research to confirm it.

197. a. *Irakezen zoeken via briefjes contact met hun familie.*
Iraqis search via little_notes contact with their family
'Iraqis are trying to establish contact with their family via little notes."
(WR-P-P-G-0000008867.head.1.s.1)

b. *Prinses Irene zoekt door het schrijven van boeken over*
princess Irene searches by the writing of books about
*natuurbeleving naar contact met en waardering van*
nature_experience to contact with and appreciation of
*'gewone mensen', (…)*
ordinary people
'Princess Irene is trying to establish contact with "ordinary people" by
writing books about experiencing nature.'
(WR-P-P-G-0000072759.p.9.s.8)

198. a. *Daarom zocht hij steun bij het volk en de soldaten.*
therefore searched he support with the people and the soldiers
'That's why he sought support from the people and the soldiers.'
(WR-P-P-G-0000005943.p.5.s.3)

b. *Topman  Worms (...) zoekt  naar steun.*
top_executive Worms  searches to  support
'Top executive Worms (...) seeks support.'
    (WR-P-P-H-0000000476.p.2.s.4)

199. a. *Honderden gezinnen, maar vooral bejaarden  zochten*
hundreds families but especially elderly_people searched
*beschutting in scholen en sporthallen.*
shelter  in schools and sports_halls
'Hundreds of families, but especially elderly people, sought shelter in schools and sports halls.'
    (WS-U-E-A-0000073325.p.1.s.8)

b. Onder druk  van (...) zochten steeds meer kleine spelers naar
under pressure of  sought ever more small players to
*beschutting in een grote(re) groep.*
shelter  in a bigger  group
'Under pressure of (...), ever more small players were seeking shelter in a big(ger) group.'
    (WR-P-P-G-0000219979.p.3.s.2)

Among the telephonic verbs, we seem to have found a fairly consistent semantic distinction, whereby the transitive construction is polysemous, being used to call animate themes and to order themes to come over, while the prepositional constituent is used to report something.

  The motoric verbs yielded a distinction in terms of conation: the transitive construction entails the successful completion of the action expressed by the verb, while the prepositional variant does not. Still, it does still depend on the verb and the theme in question whether the prepositional variant also implies that there is a serious possibility that the action will fail. This does seem to be the case for *happen* 'snap', and for *grijpen* 'grab' when it is combined with themes such as *macht* 'power' or *overwinning* 'victory', but much less so when *grijpen* 'grab' is combined with themes that are concrete items. In such cases, the prepositional variant is actually used quite often when the grabbing in question undoubtedly succeeds, as in (205) in Section 7.3 and as is evident from the second clause *en begint...* 'and starts...' in (200).

200. *Hij grijpt naar de microfoon en begint de menigte toe te spreken.*
he grabs to the microphone and starts the crowd PART to speak
'He reaches for the microphone and starts speaking to the crowd.'
    (WR-P-P-B-0000000178.p.646.s.2)

Both meaning distinctions may stem from the original directional meaning of prepositional adjuncts with the preposition *naar* 'to', from which prepositional objects have developed (cf. Subsection 2.2.1). Still, this does not mean that they are one and the same meaning distinction. For one, transitive *bellen* 'phone' or *telefoneren* 'phone' do not entail that contact with the addressee was made, as is evident from the second clause *maar ik heb niet opgenomen* 'but I didn't answer' in (201). By contrast, transitive *grijpen* 'grab', as well as the other motoric verbs, do entail such contact with the theme, and sentences such as (202) hence do not occur (cf. (134) in subsection 5.2.5.1).

201. *Hij heeft me gebeld, maar ik heb niet opgenomen.*
     he has me phoned but I have not taken_up
     'He phoned me, but I didn't answer.'
                    (WR-P-E-G-0000002332.p.665.s.1)

202. *?Hij greep      direct  zijn wapen, maar zijn holster was leeg.*
     he   immediately grabbed his  weapon but his  holster was empty
     'He immediately grabbed his weapon, but his holster was empty.'
                    (WR-P-P-B-0000000229.p.2113.s.2)

This leaves the venatic verbs *jagen* 'hunt' and *vissen* 'fish'. The manual data selection of Chapter 5 for these verbs left too few instances of either variant that were deemed interchangeable to do any serious analyses. Still, we might derive something from the selection itself. It appears that the transitive variants of these verbs are typically used in resultative sentences, while prepositional variants are used in the figurative meaning 'attempt to reach/obtain/know'.

Of course, subjective interpretation always remains a key component of the empirical research cycle (Geeraerts 2010b). This is most obviously the case for semantic differences that are not very clear-cut: for the verbs *peilen* 'gauge' and *zoeken* 'search' or even for the telephonic verbs, it is possible to interpret the results of our analyses in different ways. For instance, we have operationalized the hypothesized distinction for *zoeken* 'search' between 'seek to make/acquire' and 'look for' through, among others, the variable SEMANTIC COHERENCE TO *THE ZOEK-NAAR*-CONSTRUCTION. However, the observed positive correlation between this variable and the probability of the prepositional variant of *zoeken* 'search' can also be interpreted differently: perhaps it signals more of an aspectual distinction or a difference in terms of focus.[90] To the bottom of this, a continuous cycle of refinement and testing of semantic hypotheses is needed (Grondelaers, Speelman and Geeraerts 2008: 154–156; Geeraerts 2010b: 73–75).

---

[90] Additional tests at the level of the object, however, do indicate that the meaning difference has to be more concrete than that, as argued below.

## 9.2     Where does argument structure vary?

We can now turn to the where-question. Figure 1 in Section 3.3 sketched a constructional network of the prepositional variant, detailing the various levels of abstraction at which the alternation might function. Figure 37 retakes this network, retaining only the nodes that appear to express their own idiosyncratic meaning distinctions (cf. Perek 2015: 218).

We hope to have illustrated two points throughout this thesis with regard to the where-question. First, variation in argument structure may be determined at various levels of abstraction, and the answer to the where-question is hence dependent on the case study at hand, as argued a.o. by  Boas (2014), Perek (2015) and Pedersen (2019). This dovetails with usage-based theory, which states that one of the primary reasons that variation exists is to answer a functional need of the language user (Van de Velde 2014b). Since this functional need may manifest itself at any level of abstraction, we fully expect that variation in argument structure may also be determined at any level of abstraction. Second, the where-question can be answered empirically by investigating variation in argument structure at various levels of abstraction in a systematic way (Perek 2014).
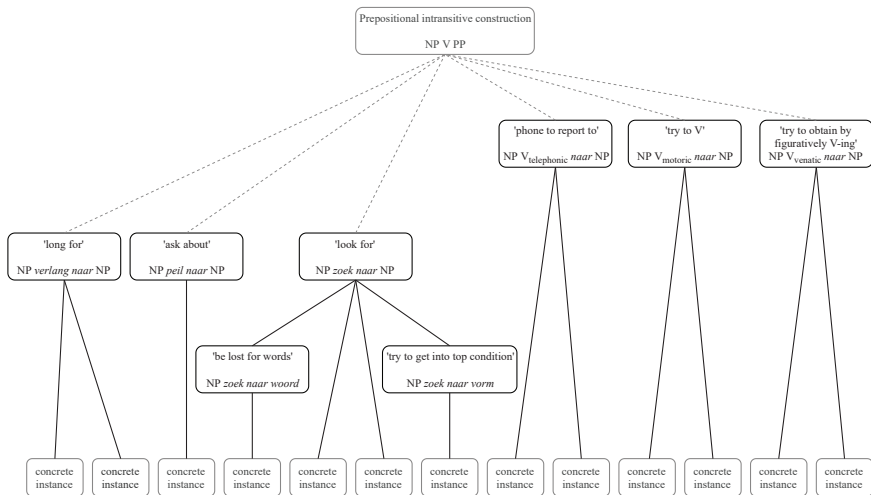


Figure 37: Pruned constructional network of the prepositional variant, retaining only the nodes necessary to account for the observed meaning distinctions.

Still, there are two important arguments in favor of the exclusive use of abstract argument constructions. First, it may be argued that each and every concrete meaning difference observed in this study is actually caused by a single, underlying abstract distinction, e.g. in terms of directionality. Indeed, the mere finding that our predictions based on the Sem1 hypothesis failed, is no argument against such an abstract distinction. For one, absence of evidence does not equal evidence of absence. For another, perhaps Sem1 interpreted the directionality distinction too literally.

Still, such an abstract distinction would suffer from the problems discussed in Section 2.3. That is, if the notion of directionality is to account for the distinction between 'demand' and 'long for' for the verb *verlangen* 'desire', as well as the distinction in terms of conation for the motoric verbs, and the distinction between literally looking for words versus trying to express a proposition for *woord zoeken* 'search word', etc., then the notion would need to be stretched to such a degree that it is no longer tenable "as a truly explanatory notion in semantics" (Lenci 2012: 14, also see Dąbrowska 2016: 483−484, 2017: 21−38 for strong arguments against the use of notions in cognitive linguistics that are vague to the point that they evade empirical falsification). This is not to argue that directionality plays no role whatsoever for the alternation, but merely that it cannot be used to explain all observed semantic contrasts.

As for high level semantic distinctions other than directionality, e.g. iconicity or affectedness, future research may indeed reveal that such a distinction is at play, either at the level of the preposition or even higher. We currently see no way, however, to conceive of how such a distinction may account for each and every semantic contrast observed in this study, unless if these distinctions would be couched in such vague terms that would make it suffer from the same problem as above.

A second argument against concrete constructions could be that the use of such constructions is non-parsimonious and should be avoided at all costs. Construction grammar in particular has been criticized for such a perceived lack of parsimony, especially since the definition of what constitutes a construction was adapted from Goldberg (1995) to Goldberg (2006) to include all fully predictable patterns that occur with sufficient frequency (Hutchinson 1974: 57−59; Culicover and Jackendoff 2005; Traugott and Trousdale 2013: 5−11).

However, the parsimony of an explanation of linguistic observations should not be measured by the number of constructions it discerns, but rather by the number of mechanisms it requires to account for its observations − at least within a usage-based framework (cf. Croft 1998).[91] In the previous section, where we tried to

---

[91] There are several other arguments in favor of the use of concrete constructions. One is based on another form of parsimony not discussed here, namely *computing parsimony* (Croft 2003: 60−62; see also Langacker 2009: 250−257). This relates to minimizing the

explain how the seemingly diverse meaning distinctions at the various levels of abstraction may have originated, we have only made reference to mechanisms that are already known in the literature and for which there is independent linguistic evidence, such as the lexical origin hypothesis and constructional contamination. Similarly, the effects of lectal and processing factors on the alternation can also be explained by independently established mechanisms, such as functional specialization and the principle of Uniform Information Density. As such, we believe we can claim that the explanations provided by this thesis are, arguably, highly parsimonious.

Note that a crucial prerequisite to come to such explanations, especially for the semantic and lectal distinctions, is the integration of lectal, semantic and processing-related effects in the study of language variation (Geeraerts 2006a: 2–6, 2006b: 30–31, 2010b). For instance, constructional contamination is fundamentally a processing effect that may lead to semantic distinctions, while functional specialization originates in the extralinguistic context, but affects both semantic and processing factors (Speelman, Grondelaers and Geeraerts 2008; Pijpops and Van de Velde 2016).

## 9.3    Alternations and allostructions

In this section, we return to the considerations regarding alternations from Section 2.1 and apply them to our case study. We first retake the various definitions of alternations and discuss how they in hindsight apply to the transitive-prepositional alternation, and then turn to the issue of the cognitive reality of the alternation factors. Finally, we discuss the notion of allostructions.

---

amount of computing required from individual language users when producing or parsing utterances. If it can be shown that maintaining a multitude of concrete constructions leads to gains in processing efficiency, such a proliferation of concrete constructions is defensible as a strategy to trade storage capacity for computing speed. While the storage capacity of the human brain is of course finite, it is in fact quite large, rendering such a trade-off attractive (cf. Bartol et al. 2015 and references cited therein). Another argument is that the redundancy entailed in employing both abstract constructions and concrete constructions is actually a desirable property of a language system, since it is a necessary requirement for any degenerative system flexible enough to adapt and evolve (Van de Velde 2014b).

## 9.3.1     Definitions of alternations

Which segments of the variation between the transitive and the prepositional variant with *naar* 'to' best fit each definition of an alternation? Regarding the psycholinguistic definition, this question was already answered in Chapter 8: the best candidates for a psycholinguistic case study, i.e. where lectal and semantic factors can best be kept under control, is the alternation with the verb *peilen* 'gauge' − but only within Belgium − and the alternation with the verb *zoeken* 'search' − but only when particular themes are excluded.

Meanwhile, the best fit for the classic or relaxed sociolinguistic definition is the alternation with the verb *peilen* 'gauge'. Here, semantic differences seem minor, meaning that the transitive and prepositional variant of this verb could, given some leeway, qualify as "alternate ways of saying 'the same' thing" (Labov 1972a: 188; Tagliamonte 2012: 2). Of course, processing-related factors would still need to be controlled for. Another alternation that might fit this definition, is the alternation of the verb *grijpen* 'grab', and probably some of its near-synonyms, like *graaien* 'grasp' and *grabbelen* 'scramble', when they are combined with concrete themes and when it is clear from the context that the action was successful (cf. (188) in Section 7.3 and (200)). We have yet to identify a lectal distinction there, however. The alternation with the verb *verlangen* 'desire' is actually also very interesting from a sociolinguistic point of view, because its seems more strictly determined in the Netherlands than in Belgium (cf. Section 6.2.1), but it clearly cannot quality as alternate ways of saying the same thing: the semantic distinction between the variants is by far too dominant.

The grammatical definition of an alternation required a systematic difference in form that corresponds to a systematic difference in meaning. The best candidates for this in the present case study are the alternations among the telephonic verbs and the motoric verbs. Here, the presence or absence of the preposition seems to trigger the same meaning difference among several verbs. For the verbs *verlangen* 'desire' and *zoeken* 'search', we also found meaning differences, but since these are limited to a particular verb, or even to particular themes, they can hardly qualify as a systematic difference.

We now turn to the definition of an alternation as choice points of the individual language user. Chapter 5 attempted to put this definition into practice for the transitive-prepositional alternation, and hence, all instances that have been retained in our dataset after the manual checking should fall under this definition − barring, most likely, a number of non-interchangeable instances that were retained in the dataset because of human error. In retrospect, however, there still remain some points of contention regarding the questions mentioned in Section 2.1: who is the individual; which are the variants; and where does a factor become too dominant?

First, for many Netherlandic language users of Dutch, the verb *peilen* 'gauge' may not constitute a choice point at all, since the prepositional variant is nearly non-existent in their country. Conversely, using the transitive variant may not be an option for some Belgian language users, given the dominance of the prepositional variant in Belgium. While *peilen* 'gauge' does constitute a choice point for the author of this thesis, there is no reason that he should be privileged as a 'representative language user' of Dutch (Geeraerts 2010b: 66–68; Verhagen 2013).

Second, other variants are often possible too, as is the case for almost every alternation (cf. Section 2.1, Van de Velde 2013: 164–165, Szmrecsanyi et al. 2016). For example, in sentence (203), the language user could opt for the transitive or prepositional variant of *zoeken* 'search', but also for *op zoek zijn* lit. 'be on search', *een zoektocht doen naar* lit. 'do a search to', *proberen te vinden* 'try to find', etc. In other instances, like (204), these variants perhaps sound somewhat less natural, but other alternatives are still possible, such as *nastreven* 'pursue', *streven naar* 'strive for', etc. For *telefoneren* 'phone', there is an additional prepositional alternative with *met* 'with' (cf. Table 2 in Section 3.2) – although this alternative is only possible if contact with the theme is made, as shown in (205) (cf. (201)).[92]

203. *Ook daar zoeken reddingswerkers nog naar vijf mensen.*
  also there search rescue_workers still to five people
  'Rescue workers are also still looking for five people over there.'
  (WR-P-P-G-0000006869.p.2.s.4)

204. *Het geloof in het Christuskind zoekt naar inzicht.*
  the faith in the Christ_child searches to insight
  'The faith in the Child Christ seeks insight.'
  (WR-P-P-G-0000119136.p.11.s.5)

205. *? Hij heeft met me getelefoneerd, maar ik heb niet opgenomen.*
  ? he has with me phoned but I have not picked_up
  'He has talked over the phone with me, but I didn't pick up.'

---

[92] The verb *bellen* 'phone' also appears with *met* 'with'. This verb-preposition combination did not make it into Table 2 in Section 3.2, however, since the Alpino-parser considers the *met*-constituent with *bellen* 'phone' to be a prepositional adjunct. The status of such constituents as either prepositional adjuncts or objects is indeed controversial (Vandeweghe and Colleman 2011; Broekhuis 2014), but the decision to consider them objects for *telefoneren* 'phone' and adjuncts for *bellen* 'phone' is still, admittedly, fairly arbitrary. Still, we expected and accepted such arbitrariness as fundamentally unavoidable (cf. Subsection 2.2.2).

Third, it could be proposed that for the verb *verlangen* 'desire', the semantic difference is too dominant, leaving no choice for the language user. The degree to which this can be argued differs from occurrence to occurrence. In (206a-b), the language does seem to have a choice whether to construe the desire more as a longing and opt for the prepositional variant, or to incline more towards a demand and use the transitive variant. In (206c), switching to the prepositional variant already seems to be harder. Similarly, it may be argued that particular lexical biases for the verb *zoeken* 'search' as in (197)-(199) are so dominant that no choice is left.

206. a.   *Mijn suggestie: ik vond het recept als basis in orde, maar*
      my  suggestion  I   found  the recipe  as  basis in  order  but
      *verlangde meer smeuïgheid en   frissigheid.*
      desired    more smoothness and  freshness
      'My suggestion: I thought the recipe was alright as a basis, but desired more smoothness and freshness.'
           (WR-P-P-G-0000025966.p.5.s.1)

    b.   *Kroatische parlementsleden      verlangden een tekst in het*
      Croatian    parliaments_members desires    a    text in  the
      *'Kroatisch', waarbij er   slechts één letter  verschil   met het*
      Croatian    whereby there only   one  letter  difference with the
      *'Bosnisch' zou    zijn.*
      Bosnian   would  be
      'Croatian members of parliament desired a text in "Croation", whereby there would be only a single letter different from "Bosnian".'
           (WR-P-P-K-0000000016.p.163.s.3)

    c.   *Bovendien verlangde de EG een  wapenstilstand als voorwaarde*
      Moreover    desired    the EG a     truce        as condition
      *voor politieke besprekingen.*
      for   political talks
      'Moreover, the EG desired a truce as condition for political talks.'
           (WR-P-P-K-0000000014.p.217.s.1)

Still, under the current definition, these three questions can be considered mere practical difficulties that require practical answers, as argued in Section 2.1. These practical answers are the following. The alternation at least presents a choice point for some language users of Dutch; the latter was considered sufficient to justify its study. To delineate the variants, we made use of formal criteria presented in Section 3.2. As for the final question, Chapter 5 details where the line between the

categorical and probabilistic was drawn: for *verlangen* 'desire', for instance, it was drawn between on the one hand inanimate and on the other animate and collective agents (cf. Subsection 5.2.2.1).

Lastly, the definition of an alternation as a practical research set-up states that the antithesis of two forms as an alternation can only be justified by an a priori hypothesis, or at least by a serious theoretical reason. The entire alternation as studied in this thesis certainly qualifies for this definition: the hypotheses presented in Chapter 4 give us sufficient reason to pit the transitive construction and the prepositional intransitive construction with *naar* against one another as variants of an alternation. In addition, the hypotheses can be used to justify why we restricted our investigation to the variants with and without *naar* 'to', thereby excluding other alternatives for *zoeken* 'search', like *op zoek zijn* lit. 'be on search' or for *verlangen* 'desire', e.g. *vragen* 'demand', *willen* 'want', *hunkeren* 'long for', etc.: the predictions of these hypotheses should still hold, even when these other variants are excluded.

## 9.3.2    Cognitive reality of alternation factors

In principle, all of the alternation factors taken up in our models could be cognitively real. However, since our analyses are based on offline corpus data, we prefer to stick to conservative conclusions in this regard. The lectal distinctions may well constitute mere differences between language users of Belgian and Netherlandic nationality, without these language users being aware of the differences (cf. Grondelaers 2016). Likewise, the tendency to place the preposition in front of complex themes when the verb precedes the theme, while omitting it in front of complex themes when the theme precedes the verb, is not necessarily registered in the brains of individual language users. It may also be a simple ad hoc side effect during the production of sentences.

While this means that these tendencies may not be cognitively real, they are still, of course, undeniably real in the sense that they are directly observable in language use. That is, they are part of grammar as "observable regularit[ies] in the language use realized by a specific community" (Geeraerts, Kristiansen and Peirsman 2010: 5; cf. also Steels 2000; Geeraerts 2010d; Van de Velde 2017). As such, they can still provide valuable insights when testing lectal and even processing-related hypotheses, as exemplified in Chapters 4, 6 and 8.

Regarding the influence of the semantic alternation factors, notably those relating to very clear distinctions, as for *verlangen* 'desire', we currently see no way to explain them other than to assume that individual language users do know about them, if only subconsciously. In other words, we see no possibility to account for these distinctions, other than to assume that they are indeed cognitively real. Still, there may well exist important differences between language users in this

regard. When faced with tendencies in language use, such as the ones observed in this thesis, language learners may come to different interpretations – just like language researchers, as mentioned above (cf. Dąbrowska 2012, and references cited therein). It may be naïve to think that all language users would agree on a single, uniquely valid distinction, especially seeing as such disagreement would unlikely lead to any serious problems in communication (Dąbrowska 2015). For instance, imagine that one language user would believe that transitive *zoeken* 'search' leans more towards 'seek to make/acquire' and prepositional *zoeken* 'search' leans more towards 'look for', while another would believe that transitive *zoeken* 'search' is more of a stationary action, while prepositional *zoeken* 'search' is more like a path towards a goal. Both would likely still be perfectly capable of communicating with one another, without this difference in beliefs ever becoming an issue, or perhaps even without it ever being noticed.

### 9.3.3    Allostructions

Allostructions are two subconstructions of a single overarching, more abstract construction, which function as the formal realizations of this more abstract construction (Cappelle 2006). The overarching construction specifies the formal and semantic properties that are shared between both allostructions, while the allostructions themselves contain the differences. Allostructions are viewed as the syntactic counterparts of allomorphs and allophones – or, more accurately, allomorphs and allophones would be considered morphological and phonological allostructions (Cappelle 2006: 21–22).

Examples are the English particle placement alternation as in (207) and the English dative alternation as in (208). There are strong arguments why the variants of these alternations should be described as allostructions; these include evidence from statistical preemption and priming experiments, where both variants have been shown to preempt and prime one another, and from sorting tasks, where language users would sort sentences of both variants together in spite of other available strategies for categorization (see the overviews in Cappelle 2006; Perek 2015: 163–167). In addition, there is the argument that when various instances do show some similarity in form and meaning, an overarching construction would be expected to arise anyway (Cappelle 2006: 21).

207. a.    *The police brought in the criminal.*
     b.    *The police brought the criminal in.*
                        (taken over from Cappelle 2006: 4)

208. a.  *Mary gave John a book.*
    b.  *Mary gave a book to John.*
                              (taken over from Perek 2012: 602)

Since we have not conducted any experiments directly aimed at this matter, we cannot say whether our variants indeed classify as allostructions. Still, if they do, it will not be the transitive and prepositional intransitive constructions at the highest level of abstraction. These two constructions hardly overlap: while Chapter 3 identified 101 verbs where they do seem to overlap, this is still only a fraction of the entire verbal inventory of both constructions, especially for the highly frequent transitive construction. Moreover, the factors determining the alternation have been shown to diverge strongly from verbal group to verbal group, from verb to verb and sometimes even from verb-theme combination to verb-theme combination.

The existence of allostructions would make more sense, however, at lower levels of abstraction. For instance, the formal and semantic overlap between the transitive *verlang*-construction and the prepositional *verlang-naar*-construction is rather obvious: both constructions contain the same verb and express a form of desire. The same holds for the other prepositional constructions and their transitive counterparts proposed in Figure 37. This would already constitute one avenue for further research: conducting experiments to determine whether these variants indeed form allostructions. The next section presents more of such possibilities.

## 9.4 Future research

As indicated in Chapter 1, the present study only aimed at providing a first exploration of the transitive-prepositional alternation, and there are many ways in which the alternation can be fruitfully further investigated. First, future studies could put other potential alternation factors under scrutiny. As for the lectal dimension, the alternation could be investigated in different corpora that allow for more fine-grained geographical and social distinctions.

Regarding semantic factors, Section 2.3 already listed the various proposals in the literature. These include distinctions in terms of aspect, affectedness, partitivity, sentience, sensuality, as well as the precision, seriousness and purpose of the action expressed by the verb. Of course, other distinctions could be relevant as well, such as the agentivity of the agent or the theme (Pijpops and Speelman 2017), the lack of intentionality (Perek 2015: 93–94), etc. Future research could focus on finding and testing operationalizations of these distinctions.

There are also many more potential processing-related factors that could still be brought to bear: e.g. other operationalizations of the complexity of the theme, the complexity of the entire sentence or of the agent, the givenness and definiteness of the agent and the theme, the influence of constructional contamination, lectal contamination, alpha- and beta-persistence (Szmrecsanyi 2005), etc. Still, our current regression models already reach excellent discrimination, with C-indexes ranging between 0.8 and 0.9 (Hosmer and Lemeshow 2000: 162).[93] In other words, while the list of potential alternation factors is virtually endless, the amount of variation left to be explained is limited.

Second, experiments could be designed to delve into the cognitive reality of these alternation factors, as in Rosenbach (2003), Klavan and Divjak (2016) and Divjak, Dąbrowska and Arppe (2016), or into the status of the variants as allostructions, as mentioned above. Third, a diachronic study is crucial for finding conclusive answers to the why-question, as argued in Section 9.1. The timing finally seems favorable for such a study, since a new diachronic corpus of Dutch is about to be released (Piersoul, Van de Velde and De Troij ms.). In such a corpus, the changes in semantics of the verbs or verb-theme combinations under scrutiny could be compared to their changes in argument structure. Fourth, the transitive-prepositional alternation still deserves an in-depth investigation of the prepositions other than *naar* 'to'. The present thesis can then function as a starting point, since we have already mapped out which prepositions alternate for which verbs.

Fifth, the constructions proposed in this thesis could be formalized. To this end, we would need a formalism that readily allows the researcher to specify constructions at various levels of abstraction. Such a formalism can be found in Fluid Construction Grammar (Steels 2011b, 2017). Further advantages of Fluid Constructions Grammar, except that it allows for various levels of abstraction, are (i) that it has been developed based on a usage-based concept of language and directly incorporates many of the principles of usage-based linguistics (Steels 2011b; van Trijp 2013; Wellens et al. 2013); and (ii) that it has consistently proven its use in theory development (Beuls 2012; Beuls and Steels 2013; van Trijp 2014; Van Eecke 2017).

Formalizing the constructions would be useful for two reasons. First, it would allow us to increase the clarity and precision of our description and formulate more precise predictions. For example, when combining the lexical *weg*-construction

---

[93] The only regression models presented in this thesis to fall below a C-index of 0.8 are the regression models at the level of the preposition, in Tables 12-13 and Tables 24-25, and the regression model at the level of the object for *woord zoeken* 'search word' in Table 27. The reason for the bad performance of the first models is that these models could not incorporate distinctions specific to each particular verb, which have been shown to be crucial in predicting the variation. The reason for the bad performance of the last model might be because important predictors are missing from it, such as an interaction between THEME COMPLEXITY and VERB-THEME ORDER.

with either the transitive *zoek*-construction or the prepositional *zoek-naar*-construction, their constructional combinations should yield the correct compositional meanings, as we find them in corpus data. Second, formalization would be the first step in building computer simulations that allow us to compare the theorized mechanisms of how constructions and their constructional meanings emerge (cf. Beuls and Steels 2013; Pijpops, Beuls and Van de Velde 2015).

There are also many avenues of future research outside of the transitive-prepositional alternation. For instance, the present study has investigated how language variation is determined differently in Belgian and Netherlandic Dutch, but its scope was of course limited to the transitive-prepositional alternation. A grand-scale study into this subject is still needed: this would reveal in how far these differences are systematic (e.g. De Troij et al. forthc.).

Many questions also remain regarding semantic differences in argument realization or language variation in general. How strong do lexical biases need to be in order to trigger the lexical origin mechanism? Which are the most common mechanisms to generate such initial lexical biases?

Finally, regarding the explanations of the Complexity Principle, we would still want to differentiate between on the one hand the production-driven account proposed in Ferreira and Dell (2000) and MacDonald (2013), and on the other the channel-driven model underlying the principle of Uniform Information Distribution (Fenk-Oczlon 2001; Jaeger 2010). Though we currently see no way of doing so using the transitive-prepositional alternation, perhaps researchers can in the future come up with clever, new operationalizations, or new case studies.

## 9.5    To conclude

To conclude this thesis, we would like to shortly summarize its contributions to various subfields of linguistics. To Dutch grammar description, this thesis provides the first systematically compiled list of the prepositions and verbs alternating between the transitive and prepositional-intransitive argument constructions. As mentioned in Chapter 3, this list has its limitations – mostly due to verbs of low frequency not being included and the unavoidable arbitrariness in delineating the prepositional object from prepositional adjuncts – but it can still form a useful basis for future research. Moreover, the investigation yielded answers to the how- and where-questions regarding the variation among verbs alternating with *naar* 'to': these constitute the descriptive findings of this thesis, summarized in Sections 9.1 and 9.2.

To the field of alternation studies, the methodological accomplishments of the thesis are perhaps more relevant. It has implemented a procedure of testing

meaning differences between two constructional variants based on linguistic theory, viz. on the lexical origin mechanism and the Principle of Semantic Coherence. This procedure consists of a single work flow that of integrates collostructional analyses, distributional vectors and mixed effects regression modelling. In addition, it has illustrated how Memory-based Learning may be employed as a scalable technique for hypothesis-generation at low levels of abstraction.

The results of respectively the lectal and complexity hypotheses would be most relevant to sociolinguistics and psycholinguistics: these results provide additional evidence for functional specialization due to delayed standardization, and for production- and channel-based explanations of the Complexity Principle.

Meanwhile, cognitive linguists would be most interested in the integration of the study of social and processing effects with the study of semantic distinctions (Geeraerts 2010b). Here, we have shown how the lexical origin mechanism may function at a low level of abstraction, viz. for argument constructions where the verb slot is fixed. Moreover, we have provided additional evidence for this lexical origin mechanism, and for the claim that argument structure can be determined at various levels of abstraction.

Finally, to all subfields of linguistics, we hope that this thesis has illustrated the usefulness of asking big how-, why- and especially where-questions when studying language, even when dealing with only a modest case study.

# References

2007. The British National Corpus, version 3 (BNC XML Edition). University of Oxford: Bodleian Libraries, on behalf of the BNC Consortium.

2014. Oxford English dictionary. Oxford University Press.

Absillis, Kevin, Jürgen Jaspers & Sarah Van Hoof (eds.). 2012. *De manke usurpator: over Verkavelingsvlaams*. Ghent: Academia press.

Abraham, Werner and Charl Jacobus Conradie. 2001. *Präteritumschwund und Diskursgrammatik.* Amsterdam: Benjamins.

Adler, Julia. 2011. Dative alternations in German. The argument realization options of transfer verbs. Dissertation Hebrew University.

Altenberg, Bengt. 1998. On the phraseology of spoken English: The evidence of recurrent word-combinations. In Anthony Paul Cowie (ed.), *Phraseology: theory, analysis, and applications*, 101–122. Oxford: Clarendon Press.

Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert and Arne Zeschel. 2010. Cognitive Corpus Linguistics: Five Points of Debate on Current Theory and Methodology. *Corpora: Corpus-Based Language Learning, Language Processing and Linguistics* 5(1). 1–27.

Arppe, Antti. 2008. Univariate, bivariate and multivariate methods in corpus-based lexicography. Dissertation University of Helsinki.

Arppe, Antti. 2009. Linguistic choices vs. probabilities. How much and what can linguistic theory explain? In Sam Featherston & Susanne Winkler (eds.), *The Fruits of Empirical Linguistics. Vol. 1 Process*, 1–14. Berlin/New York: Mouton De Gruyter.

Augustinus, Liesbeth. 2015. Complement Raising and Cluster Formation in Dutch. A Treebank-supported Investigation.

Balk-Smit Duyzentkunst, Frida. 2000. *Grammatica van het Nederlands [Grammar of Dutch]*. Den Haag: Sdu.

Bartol, Thomas, Cailey Bromer, Justin Kinney, Michael Chirillo, Jennifer Bourne, Kristen Harris and Terrence Sejnowski. 2015. Nanoconnectomic upper bound on the variability of synaptic plasticity. *eLife* 4. e10778.

Bates, Douglas, Martin Mächler, Ben Bolker and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software* 67(1). 1–48.

Beavers, John. 2011. On affectedness. *Natural Language & Linguistic Theory* 29(2). Dordrecht. 335–370.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten Christiansen, William Croft, Nick Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman and Tom Schoenemann. 2009. Language Is a Complex Adaptive System: Position Paper. *Language Learning* 59(1). Malden, USA. 1–26.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. Walter de Gruyter. 110–142.

Bentz, Christian, Annemarie Verkerk, Douwe Kiela, Felix Hill and Paula Buttery. 2015. Adaptive Communication: Languages with More Non-Native Speakers Tend to Have Fewer Word Forms. *PLoS ONE* 10(6). Public Library of Science (PLoS).

Bernaisch, Tobias, Stefan Thomas Gries and Joybrato Mukherjee. 2014. The Dative Alternation in South Asian English(es): Modelling Predictors and Predicting Prototypes. *English World-Wide: A Journal of Varieties of English* 35(1). 7–31.

Bernolet, Sarah and Timothy Colleman. 2016. Sense-based and lexeme-based alternation biases in the Dutch dative alternation. In Jiyoung Yoon & Stefan Thomas Gries (eds.), *Corpus-based Approaches to Construction Grammar*, 165–198. Amsterdam/Philadelphia: John Benjamins.

Beuls, Katrien. 2012. Inflectional patterns as constructions: Spanish verb morphology in Fluid Construction Grammar. *Constructions and Frames* 4(2). 231–252.

Beuls, Katrien and Luc Steels. 2013. Agent-Based Models of Strategies for the Emergence and Evolution of Grammatical Agreement. *PLoS ONE* 8(3). e58960.

Bloem, Jelke, Arjen Versloot and Fred Weerman. 2017. Verbal cluster order and processing complexity. *Language Sciences* 60. 94–119.

Boas, Hans. 2008. Resolving form-meaning discrepancies in construction grammar. In Jaakko Leino (ed.), *Constructional reorganization*, 11–36. Amsterdam: John Benjamins.

Boas, Hans. 2010. The syntax-lexicon continuum in Construction Grammar. A Case study of English communication verbs. *Belgian Journal of Linguistics* 24(1). 54–82.

Boas, Hans. 2014. Lexical and phrasal approaches to argument structure: Two sides of the same coin. *Theoretical Linguistics* 40(1–2). 89–112.

Bock, Kathryn, David Irwin and Douglas Davidson. 2004. Putting First Things First. *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, 224–250. New York: Psychology Press.

Bolinger, Dwight. 1972. *That's that*. The Hague: Mouton.

Bolinger, Dwight. 1980. Wanna and the Gradience of Auxiliaries. In Günter Brettschneider & Christian Lehmann (eds.), *Wege zur Universalienforschung: sprachwissenschaftliche Beiträge zum 60. Geburtstag von Hansjakob Seiler*, 292–299. Tübingen: Gunter Narr.

Booij, Geert. 2009. A Constructional Analysis of Quasi-Incorporation in Dutch. *Gengo Kenkyu* 135. 5–28.

Bouma, Gosse and Geert Kloosterman. 2002. Querying Dependency Treebanks in XML. *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*, 1686–1691.

Bouma, Gosse and Geert Kloosterman. 2007. Mining Syntactically Annotated Corpora with XQuery. *Proceedings of the Linguistics Annotation Workshop (ACL 07)*, 17–24.

Bouma, Gosse. 2017. Om-omission. In Martijn Wieling, Martin Kroon, Gertjan van Noord & Gosse Bouma (eds.), *From Semantics to Dialectometry*, 65–73. Groningen: College Publications.

Boyd, Jeremy, Erin Gottschalk and Adele Eva Goldberg. 2009. Linking Rule Acquisition in Novel Phrasal Constructions. *Language Learning* 59(s1). Malden, USA: Blackwell Publishing Inc. 64–89.

Brennan, Susan and Maurice Williams. 1995. The feeling of another's knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language* 34(3). 383–398.

Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in Search of Its Evidential Base*, 77–96. Berlin: Mouton De Gruyter.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina and Rolf Harald Baayen. 2007. Predicting the dative alternation. In Gerolf Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive Foundations of Interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan and Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua* 118(2). 245–259.

Bresnan, Joan and Marilyn Ford. 2010. Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English. *Language: Journal of the Linguistic Society of America* 86(1). 168–213.

Broccias, Cristiano. 2001. Allative and ablative at-constructions. In Mary Adronis, Christopher Ball, Elston Heide & Sylvain Neuvel (eds.), *CLS 37: The Main Session. Papers from the 37th Meeting of the Chicago Linguistic Society*, 67–82. Chicago: Chicago Linguistic Society.

Broekhuis, Hans. 2004. Het voorzetselvoorwerp. *Nederlandse Taalkunde* 9. 31–97.

Broekhuis, Hans. 2007. Worstelen met het voorzetselvoorwerp: opmerkingen bij Ina Schermer-Vermeers artikel. *Nederlandse Taalkunde* 12(4). 351–358.

Broekhuis, Hans. 2014. Dubbel-voorzetselobjectconstructies? In Freek Van de Velde, Hans Smessaert, Frank Van Eynde & Sara Verbrugge (eds.), *Patroon en argument. Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst*, 103–113. Universitaire Pers Leuven.

Broekhuis, Hans, Norbert Corver and Riet Vos. 2013. *Syntax of Dutch. Verbs and Verb Phrases. Volume 1*. Amsterdam: Amsterdam University Press.

Broekhuis, Hans and Norbert Corver. 2013. *Syntax of Dutch. Verbs and Verb Phrases. Volume 2*. Amsterdam: Amsterdam University Press.

Bürkle, Daniel. 2015. The acquisition of sentence alternations: How children understand and use the English dative alternation. University of Canterbury.

Bybee, Joan. 2006. From Usage to Grammar: The Mind's Response to Repetition. *Language* 82(4). 711–733.

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Cappelle, Bert. 2006. Particle placement and the case for "allostructions." In D. Schönefeld (ed.), *Constructions Special Volume 1 — Constructions all over: case studies and theoretical implications*, vol. 4.

Carroll, Ryan, Ragnar Svare and Joseph Salmons. 2012. Quantifying the evolutionary dynamics of German verbs. *Journal of Historical Linguistics* 2(2). 153–172.

Casenhiser, Devin and Adele Eva Goldberg. 2005. Fast mapping between a phrasal form and meaning. *Developmental Science* 8(6). Oxford, UK: Blackwell Publishing Ltd. 500–508.

Chandler, Tertius. 1987. *Four thousand years of urban growth*. Lampeter: Edwin Mellen Press.

Cheshire, Jenny. 1987. Syntactic variation, the linguistic variable, and sociolinguistic theory. *Linguistics*.

Claes, Jeroen. 2015. Competing constructions: The pluralization of presentational haber in Dominican Spanish. *Cognitive Linguistics* 26(1). Walter de Gruyter GmbH. 1–30.

Clark, Eva. 1978. Discovering what words can do. *Papers from the Parasession on the Lexicon*, 34–57.

Clark, Herbert. 1996. *Using language*. Cambridge: Cambridge University Press.

Clark, Herbert. 2004. Pragmatics of Language Performance. In Laurence Robert Horn & Gregory Ward (eds.), *The Handbook of Pragmatics*, 365–382. Walden: Blackwell.

Clark, Herbert and Gregory Murphy. 1982. Audience Design in Meaning and Reference. *Advances in Psychology* 9. 287–299.

Clark, Herbert and Jean Fox Tree. 2002. Using uh and um in spontaneous speaking. *Cognition* 84(1). Elsevier B.V. 73–111.

Clyne, Michael (ed.). 1992. *Pluricentric languages: differing norms in different nations*. (Contributions to the Sociology of Language 62). Berlin: de Gruyter.

Collard, Phillip, Martin Corley, Lucy MacGregor and David Donaldson. 2008. Attention Orienting Effects of Hesitations in Speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34(3). 696–702.

Colleman, Timothy. 2006. De Nederlandse datiefalternantie. Een constructioneel en corpusgebaseerd onderzoek. Dissertation Ghent Univerisity.

Colleman, Timothy. 2009a. Verb disposition in argument structure alternations: a corpus study of the dative alternation in Dutch. *Language Sciences* 31(5). 593–611.

Colleman, Timothy. 2009b. The Semantic Range of the Dutch Double Object Construction: A Collostructional Perspective. *Constructions and Frames* 1(2). 190–221.

Colleman, Timothy. 2010. Beyond the dative alternation: The semantics of the Dutch aan-Dative. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, 271–304. Berlin/Boston: Mouton De Gruyter.

Colleman, Timothy. 2014. Nog eens de dubbelvoorzetselobjectconstructie. In Freek Van de Velde, Hans Smessaert, Frank Van Eynde & Sara Verbrugge (eds.), *Patroon en argument. Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst*, 115–127. Leuven: Universitaire Pers Leuven.

Colleman, Timothy and Bernard De Clerck. 2009. "Caused motion"? The semantics of the English to-dative and the Dutch aan-dative. *Cognitive Linguistics* 20(1). 5–42.

Colleman, Timothy and Bernard De Clerck. 2011. Constructional semantics on the move: On semantic specialization in the English double object construction. *Cognitive Linguistics* 22(1). 183.

Colleman, Timothy and Martine Delorge. 2010. Inhoudsobjecten, partnerobjecten, en de dubbel-voorzetselobject-constructie. In Johan De Caluwe & Jacques Van Keymeulen (eds.), *Voor Magda: artikelen voor Magda Devos bij haar afscheid van de Universiteit Gent*, 111–126. Ghent: Academic Press.

Collins, Michael Xavier. 2014. Information Density and Dependency Length as Complementary Cognitive Models. *Journal of Psycholinguistic Research* 43(5). 651–681.

Condamines, Anne. 2018. Is "to fish in a river" equivalent to "to fish a river"? *Cognitive Linguistic Studies* 5(2). 208–229.

Corley, Martin and Robert Hartsuiker. 2003. Hesitation in speech can um help a listener understand. In Richard Alterman & David Kirsh (eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, 276–281. Boston: Cognitive Science Society.

Coupé, Christophe, Yoon Mi Oh, Dan Dediu and François Pellegrino. 2019. Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances* 5(9). eaaw2594.

Cover, Thomas and Joy Thomas. 1991. *Elements of information theory*. Hoboken: Wiley-Interscience.

Croft, William. 1998. Linguistic evidence and mental representations. *Cognitive Linguistics* 9(2). 151–173.

Croft, William. 2001. *Radical construction grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, William. 2003. Lexical rules vs. constructions. A false dichotomy. In Hubert Cuyckens, Thomas Berg, René Dirven & Klaus-Uwe Panther (eds.), *Motivation in language : studies in honor of Günter Radden*, 49–68. Stanford: CSLI Publications.

Culicover, Peter and Ray Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford university press.

D'Arcy, Alexandra. 2014. Functional Partitioning and Possible Limits on Variability: A View of Adjective Comparison from the Vernacular. *Journal of English Linguistics* 42(3). Los Angeles, CA: SAGE Publications. 218–244.

Dąbrowska, Ewa. 2006. Low-level schemas or general rules? The role of diminutives in the acquisition of Polish case inflections. *Language Sciences* 28(1). Elsevier B.V.

Dąbrowska, Ewa. 2012. Different Speakers, Different Grammars: Individual Differences in Native Language Attainment. *Linguistic Approaches to Bilingualism* 2(3). 219–253.

Dąbrowska, Ewa. 2014. Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics* 25(4). 617–653.

Dąbrowska, Ewa. 2015a. Individual differences in grammatical knowledge. *Handbook of Cognitive Linguistics*, vol. 39, 650–668. Walter de Gruyter GmbH.

Dąbrowska, Ewa. 2015b. Language in the mind and in the community. In Jocelyne Daems, Eline Zenner, Kris Heylen & Dirk Speelman (eds.), *Change of Paradigms – New Paradoxes. Recontextualizing Language and Linguistics*, 221–235. Berlin: De Gruyter.

Dąbrowska, Ewa. 2016. Cognitive Linguistics' seven deadly sins. *Cognitive Linguistics* 27(4). 479–491.

Dąbrowska, Ewa. 2017. *Ten lectures on grammar in the mind*. (Distinguished Lectures in Cognitive Linguistics). Leiden: Brill.

Daelemans, Walter and Antal van den Bosch. 2005. *Memory-based language processing*. Cambridge: Cambridge university press, 2005.

Daelemans, Walter, Jakub Zavrel, Ko van der Sloot and Antal van den Bosch. 2010. TiMBL: Tilburg Memory-Based Learner Reference Guide. Tilburg.

Daems, Jocelyne, Kris Heylen and Dirk Geeraerts. 2015. Wat dragen we vandaag: een hemd met blazer of een shirt met jasje? *Taal en Tongval* 67(2). Amsterdam: Amsterdam University Press. 307–342.

Dal, Ingerid. 1962. *Kurze Deutsche syntax auf historischer grundlage*. Tübingen: Niemeyer.

Dale, Rick and Gary Lupyan. 2012. Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems* 15(3). 1150017-1-1150017–16.

Davies, Mark. 2013. Corpus of News on the Web (NOW): 3+ billion words from 20 countries, updated every day.

De Caluwe, Johan, Steven Delarue, Anne-Sophie Ghyselen & Chloé Lybaert (eds.). 2013. *Tussentaal: over de talige ruimte tussen dialect en standaardtaal in Vlaanderen: inleiding*. (Studia Germanica Gandensia Libri). Academia Press.

De Caluwe, Johan. 2017. Van AN naar BN, NN, SN... Het Nederlands als pluricentrische taal. In Gert De Sutter (ed.), *De vele gezichten van het Nederlands in Vlaanderen: Een inleiding tot de variatietaalkunde, eerste druk*, 121–141. Leuven: Acco.

de Bruin, Cebus Cornelis (ed.). 1970. *Het Luikse diatesseron*. Leiden: E.J. Brill.

De Grauwe, Luc. 2014. Noemen voor "noemen" en "heten" (Frans nommer/se nommer). Een polycausaal verhaal over opkomst en verspreiding van een "niet-algemeen Belgisch-Nederlands" woordgebruik. *Bulletin de la commission royale de toponymie & dialectologie/ Handelingen van de koninklijke commissie voor toponymie en dialectologie [Acts of the royal commission for toponomy and dialectology]* 86. Peeter. 73–125.

De Greve, Willem. 2009. De problematische status van het voorzetselvoorwerp in het moderne Nederlands. Een kritisch literatuuronderzoek. Master's thesis, UGent.

den Boon, Ton & Dirk Geeraerts (eds.). 2005. *Van Dale Groot woordenboek van de Nederlandse taal*. 14th ed. Antwerpen/Utrecht: Van Dale Lexicography.

den Hertog, Cornelis Herman. 1903. *Nederlandsche spraakkunst : handleiding ten dienste van aanstaande (taal)onderwijzers*. Amsterdam: Versluys.

De Pascale, Stefano. 2019. Token-based vector space models as semantic control in lexical sociolectometry. Dissertation University of Leuven.

der Gucht, Fieke, Jessie De Caluwé, Mathilde Jansen and Nicoline van der Sijs. 2018. *Atlas van de Nederlandse taal: Editie Vlaanderen*. Tielt: Lannoo.

De Smet, Isabeau, Katrien Beuls, Dirk Pijpops and Freek Van de Velde. 2017. Language-specific differences in regularization rates of the Germanic preterite. *International Conference on Historical Linguistics (ICHL)*. 2 August, San Antonio.

De Smet, Isabeau and Freek Van de Velde. 2019. Reassessing the evolution of West Germanic preterite inflection. *Diachronica* 36(2).

de Swart, Peter. 2014. Prepositional Inanimates in Dutch: A Paradigmatic Case of Differential Object Marking. *Linguistics* 52(2). 445–468.

De Troij, Robbert, Stefan Grondelaers, Dirk Speelman and Antal van den Bosch. Lexicon or grammar? Using memory-based learning to investigate the syntactic relationship between Netherlandic and Belgian Dutch. Forthcoming. *Natural Language Engineering*.

De Vaere, Hilde, Ludovic De Cuypere and Klaas Willems. 2018. Alternating constructions with ditransitive geben in present-day German. *Corpus Linguistics and Linguistic Theory*. Mouton De Gruyter. Published online, ahead of print.

de Vriendt, Sera. 1965. *Sterke werkwoorden en sterke werkwoordsvormen in de 16e eeuw*. Brussel: Belgisch interuniversitair centrum voor neerlandistiek.

Diessel, Holger. 2015. Usage-based construction grammar. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handboek of Cognitive Linguistics*, 296–322. Berlin: Mouton De Gruyter.

Diessel, Holger. 2017. Usage-Based Linguistics. In Mark Aronoff (ed.), *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.

Divjak, Dagmar, Ewa Dąbrowska and Antti Arppe. 2016. Machine Meets Man: Evaluating the Psychological Reality of Corpus-Based Probabilistic Models. *Cognitive Linguistics* 27(1). 1–33.

Dixon, Robert M W. 1991. *A new approach to English grammar, on semantic principles*. Oxford: Clarendon.

Dodson, Kelly and Michael Tomasello. 1998. Acquiring the Transitive Construction in English: The Role of Animacy and Pronouns. *Journal of Child Language* 25(3). 605–622.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Driessen, Geert. 2005. In Dutch? Usage of Dutch Regional Languages and Dialects. *Language, Culture and Curriculum* 18(3).

Drinka, Bridget. 2004. Präteritumschwund: Evidence for areal diffusion. In Werner Abraham (ed.), *Focus on Germanic Typology*, 211–240. Berlin: Akademie Verlag.

Dubois, Tanguy. Forthcoming. The Complexity Principle and Lexical Complexity in the English and Dutch Dative Alternation. In Melanie Röthlisberger, Eva Zehentner & Timothy Colleman (eds.), *Ditransitive constructions in Germanic languages: Diachronic and synchronic aspects*. Amsterdam: John Benjamins.

Duden. 2009. *Die Grammatik.* 8th edn. Mannheim: Dudenverlag.

Duinhoven, Antonius Maria. 1989. Het voorzetselvoorwerp. Een zinspatroon in wording. *De nieuwe taalgids* 82. 40–55.

Ellis, Nick and Fernando Ferreira-Junior. 2009. Construction Learning as a Function of Frequency, Frequency Distribution, and Function. *Modern Language Journal* 93(3). 370–385.

Ellis, Nick and Matthew Brook O'Donnell. 2011. Robust language acquisition - an emergent consequence of language as a complex adaptive system. *Proceedings of the 33rd annual conference of the cognitive science society*, 3512–3517. Austin: Cognitive Science Society.

Ellis, Nick and Matthew Brook O'Donnell. 2012. Statistical construction learning. Does a Zipfian problem space ensure robust language learning? In Patrick Rebuschat & John Williams (eds.), *Statistical learning and language acquisition*, 265–304. Berlin: Mouton De Gruyter.

Elsness, Johan. 1984. That or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies* 65. 519–368.

Erman, Britt and Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text* 20(1). 29–62.

Fenk, August and Gertraud Fenk-Oczlon. 1993. Menzerath's law and the constant flow of linguistic information. In Reinhard Köhler & Burghard Rieger (eds.), *Contributions to quantitative linguistics*, 11–31. Dordrecht: Kluwer Academic Publishers.

Fenk-Oczlon, Gertraud. 2001. Familiarity, information flow, and linguistic form. In Joan Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure [Typological Studies in Language 45]*, 431–448. Amsterdam: John Benjamins.

Ferreira, Victor and Elizabeth Schotter. 2013. Do verb bias effects on sentence production reflect sensitivity to comprehension or production factors? *The Quarterly Journal of Experimental Psychology* 66(8). Routledge. 1548–1571.

Ferreira, Victor and Gary Dell. 2000. Effect of Ambiguity and Lexical Availability on Syntactic and Lexical Production. *Cognitive Psychology* 40(4). 296–340.

Ferreira, Victor and Melanie Hudson. 2011. Saying "that" in Dialogue: The Influence of Accessibility and Social Factors on Syntactic Production. *Language and Cognitive Processes* 26(10). 1736–1762.

Fonteyn, Lauren and Andrea Nini. 2018. Investigating the ing-form network in the idiolects of 17th century authors. *International Conference on Construction Grammar (ICCG-10)*. 17 July, Paris.

Fox Tree, Jean and Herbert Clark. 1997. Pronouncing "the" as "thee" to signal problems in speaking. *Cognition* 62(2). 151–167.

Fox, John and Sanford Weisberg. 2011. *An R companion to applied regression*. 2nd edn. Los Angeles: Sage.

Fox, John. 2003. Effect displays in R for generalised linear models. *Journal of Statistical Software* 8. 1–27.

Garnsey, Susan, Neal Pearlmutter, Elizabeth Myers and Melanie Lotocky. 1997. The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language* 37(1). 58–93.

Geeraerts, Dirk. 1988. Cognitive grammar and the history of lexical semantics. In Brygida Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*, 647–678. Amsterdam/Philadelphia: John Benjamins.

Geeraerts, Dirk. 1993. Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics* 4(3). 223–272.

Geeraerts, Dirk. 2001. Een zondagspak? Het Nederlands in Vlaanderen: gedrag, beleid, attitudes. *Ons erfdeel* 44. 337–344.

Geeraerts, Dirk. 2005. Lectal variation and empirical data in Cognitive Linguistics. In Francisco Ruiz de Mendoza Ibáñez & Sandra Peña Cervel (eds.), *Cognitive Linguistics. Internal Dynamics and Interdisciplinary Interactions*, 163–189. Berlin/New York: Mouton De Gruyter.

Geeraerts, Dirk. 2006a. Introduction. A rough guide to Cognitive Linguistics. In Dirk Geeraerts (ed.), *Cognitive Linguistics. Basic Readings*, 1–28. Berlin: Mouton De Gruyter.

Geeraerts, Dirk. 2006b. Methodology in Cognitive Linguistics. In Gitte Kristiansen, Michel Achard, René Dirven & Francisco Ruiz de Mendoza Ibañez (eds.), *Cognitive Linguistics: Current Applications and Future Perspectives*, 21–49. (Applications of Cognitive Linguistics [ACL]). Berlin/New York: Mouton De Gruyter.

Geeraerts, Dirk (ed.). 2006c. *Cognitive linguistics: basic readings*. (Cognitive Linguistics Research 34). Berlin: Mouton De Gruyter.

Geeraerts, Dirk. 2010a. Recontextualizing grammar: Underlying trends in thirty years of Cognitive Linguistics. In Elzbieta Tabakowska, Michal Choinski & Lukasz Wiraszka (eds.), *Cognitive Linguistics in Action: From Theory to Application and Back*, 71–102. (Applications of Cognitive Linguistics [ACL]). Berlin/New York: Mouton De Gruyter.

Geeraerts, Dirk. 2010b. The doctor and the semantician. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, 63–78. Mouton De Gruyter; Berlin.

Geeraerts, Dirk. 2010c. *Theories of lexical semantics*. (Oxford Linguistics). Oxford: Oxford university press.

Geeraerts, Dirk. 2010d. Schmidt redux: How systematic is the linguistic system if variation is rampant? In Kasper Boye & Elisabeth Engberg-Pedersen (eds.), *Language Usage and Language Structure*, 237–262. (Trends in Linguistics. Studies and Monographs [TiLSM]). Berlin/New York: Mouton De Gruyter.

Geeraerts, Dirk. 2017. Het kegelspel der taal. De naoorlogse evolutie van de Standaardnederlandsen. In Gert De Sutter (ed.), *De vele gezichten van het Nederlands in Vlaanderen. Een inleiding tot de variatietaalkunde.*, 100–120. Leuven: Acco Uitgeverij.

Geeraerts, Dirk, Stefan Grondelaers and Dirk Speelman. 1999. *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding- en voetbaltermen*. Amsterdam: P. J. Meertens-Instituut.

Geeraerts, Dirk, Gitte Kristiansen and Yves Peirsman. 2010. *Advances in cognitive sociolinguistics*. (Ed.) Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman. *Advances in Cognitive Sociolinguistics*. (Cognitive Linguistics Research [CLR]). Berlin/New York: Mouton De Gruyter.

Geeraerts, Dirk and Gitte Kristiansen. 2015. Variationist linguistics. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 366–388. Berlin: Mouton De Gruyter.

Geleyn, Tim and Timothy Colleman. 2015. De aan-constructie in het 17de-eeuwse Nederlands. *Taal en Tongval* 67(2). 211–245.

Geleyn, Tim. 2017. Syntactic variation and diachrony The case of the Dutch dative alternation. *Corpus Linguistics and Linguistic Theory* 13(1). Berlin: Walter de Gruyter GmbH. 65–96.

Gennari, Silvia and Maryellen Macdonald. 2009. Linking production and comprehension processes: The case of relative clauses. *Cognition* 111(1). 1–23.

Gerritsen, Willem Pieter & Willem Wilmink (eds.). 1994. *De reis van Sint Brandaan: een reisverhaal uit de twaalfde eeuw*. 6de dr. Amsterdam: Ooievaar.

Ghyselen, Anne-Sophie. 2016. Verticale structuur en dynamiek van het gesproken Nederlands in Vlaanderen: een empirische studie in Ieper, Gent en Antwerpen. Dissertation Ghent University.

Ghyselen, Anne-Sophie and Jacques van Keymeulen. 2014. Dialectcompetentie en functionaliteit van het dialect in Vlaanderen anno 2013. *Tijdschrift voor Nederlandse Taal- en Letterkunde* 130(2). 117–139.

Gilquin, Gaëtanelle and Stefan Thomas Gries. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5(1). 1–26.

Gleitman, Lila, David January, Rebecca Nappa and John Trueswell. 2007. On the Give and Take between Event Apprehension and Utterance Formulation. *Journal of Memory and Language* 57(4). 544–569.

Glynn, Dylan. 2010. Testing the hypothesis. Objectivity and verification in usage-based Cognitive Semantics. In Dylan Glynn & Kerstin Fisher (eds.), *Corpus-Driven Cognitive Semantics. Quantitaive approaches*, 239–270. Berlin: Mouton De Gruyter.

Glynn, Dylan & Kerstin Fischer (eds.). 2010. *Quantitative methods in cognitive semantics: corpus-driven approaches*. (Cognitive Linguistics Research 46). Berlin: Mouton De Gruyter.

Glynn, Dylan & Justyna Robinson (eds.). 2014. *Corpus Methods for Semantics Quantitative studies in polysemy and synonymy*. Amsterdamμ: John Benjamins.

Gochet, Alexis-Marie. 1886. *Atlas de géographie physique, politique et historique à l'usage de l'enseignement primaire et de l'enseignement moyen [Atlas of physical, political and historical geography for employment in primary and secundary education]*. Alost: Procure des frères.

Goeman, Ton and Willy Jongenburger. 2009. Dimensions and determinants of dialect use in the Netherlands at the individual and regional levels at the end of the twentieth century. *International Journal of the Sociology of Language* 196–197. 31–72.

Goldberg, Adele Eva. 1992. The inherent semantics of argument structure: the case of the English ditransitive. *Cognitive Linguistics* 3. 37–74.

Goldberg, Adele Eva. 1995. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago press.

Goldberg, Adele Eva. 1999. The emergence of the semantics of argument structure constructions. In Brian Macwhinney (ed.), *Emergence of Language*, 197–212. Hillsdale: Lawrence Earlbaum Associates.

Goldberg, Adele Eva. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*.

Goldberg, Adele Eva. 2006. *Constructions at work: the nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, Adele Eva. 2013. Argument Structure Constructions versus Lexical Rules or Derivational Verb Templates. *Mind & Language* 28(4). 435–465.

Goldberg, Adele Eva, Devin Casenhiser and Nitya Sethuraman. 2004. Learning Argument Structure Generalizations. *Cognitive Linguistics* 15(3). 289–316.

Gramley, Stephan and Kurt-Michael Pätzold. 2004. *A survey of modern English*. 2nd ed. London: Routledge.

Green, Georgia. 1974. *Semantics and syntactic regularity*. Bloomington: Indiana university press.

Gries, Stefan Thomas. 2002. The influence of processing on grammatical variation: particle placement in English. In Nicole Dehé, Ray Jackendoff, Andrew McIntyre & Silke Urban (eds.), *Verb-particle explorations*, 169–288. Berlin/New York: Mouton De Gruyter.

Gries, Stefan Thomas. 2003. *Multifactorial analysis in corpus linguistics: a study of particle placement*. New York: Continuum.

Gries, Stefan Thomas. 2007. Coll.analysis 3.2a.

Gries, Stefan Thomas. 2012. Frequencies, probabilities, and association measures in usage-/examplar-based linguistics. Some necessary clarifications. *Studies in Language* 36(3). 477–510.

Gries, Stefan Thomas. 2013. *Statistics for linguistics with R. A practical introduction*. 2nd edn. Berlin: De Gruyter.

Gries, Stefan Thomas and Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on "alternations." *International journal of corpus linguistics* 9(1). 97–130.

Grondelaers, Stefan. 2000. De distributie van niet-anaforisch er buiten de eerste zinsplaats: sociolexicologische, functionele en psycholinguïstische aspecten van er's status als presentatief signaal. Dissertation University of Leuven.

Grondelaers, Stefan. 2013. Attitude measurements in the Low Countries. In Frans Hinskens & Johan Taeldeman (eds.), *Language and Space: Dutch*, 586–602. Berlin: Mouton De Gruyter.

Grondelaers, Stefan. 2016. Subtitles, tweets, and syntax. Big data from small input. *New ways of analyzing syntactic variation*. Ghent: 20 May.

Grondelaers, Stefan, Katrien Deygers, Hilde Van Aken, Vicky Van den Heede and Dirk Speelman. 2000. Het CONDIV-corpus geschreven Nederlands [The CONDIV-corpus of written Dutch]. *Nederlandse Taalkunde* 5(4). 356–363.

Grondelaers, Stef, Dirk Speelman and Dirk Geeraerts. 2002. Regressing on "er". Statistical analysis of texts and language variation. In Annie Morin & Pascale Sébillot (eds.), *6th International Conference on the Statistical Analysis of Textual Data*, 335–346. Rennes: Institut National de Recherche en Informatique et en Automatique.

Grondelaers, Stefan, Dirk Speelman and Dirk Geeraerts. 2003. De distributie van er in het gesproken Nederlands. *Paper presented at the workshop spraakmakende spraak*. Nijmegen.

Grondelaers, Stefan, Dirk Geeraerts and Dirk Speelman. 2007. A case for cognitive corpus linguistics. In Monica Gonzalez-Marquez, Irene Mittelberg, Seana Coulson & Michael Spivey (eds.), *Methods in Cognitive Linguistics*, 149–169. (Human Cognitive Processing [HCP]). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Grondelaers, Stefan and Dirk Speelman. 2007. Regression as a referee: a theoretical and practical way out of the syntactic-variation impasse in sociolinguistics. 2007.

Grondelaers, Stefan and Dirk Speelman. 2008. Constructional near-synonymy, individual variation, grammaticality judgements. Can careful design and participant ignorance overcome the ill reputation of questionnaires. *Cognitive and Functional Perspectives on Dynamix Tendencies in Languages*. Tartu.

Grondelaers, Stefan, Dirk Speelman and Dirk Geeraerts. 2008. National variation in the use of er "there". Regional and diachronic constraints on cognitive explanations. In Gitte Kristiansen & René Dirven (eds.), *Cognitive Sociolinguistics. Language Variation, Cultural Models, Social Systems*, 153–203. (Cognitive Linguistics Research [CLR]). Berlin/New York: Mouton De Gruyter.

Grondelaers, Stefan, Dirk Speelman, Denis Drieghe, Marc Brysbaert and Dirk Geeraerts. 2009. Introducing a new entity into discourse: Comprehension and production evidence for the status of Dutch er "there" as a higher-level expectancy monitor. *Acta Psychologica* 130(2). 153–160.

Grondelaers, Stefan and Roeland van Hout. 2010. Is Standard Dutch with a regional accent standard or not? Evidence from native speakers' attitudes. *Language Variation and Change* 22. 221–239.

Grondelaers, Stefan and Roeland van Hout. 2011. The Standard Language Situation in the Low Countries: Top-Down and Bottom-Up Variations on a Diglossic Theme. *Journal of Germanic Linguistics* 23(3). 199–243.

Grondelaers, Stefan, Antal van den Bosch, Dirk Speelman and Roeland van Hout. 2015. Comparing memory-based learning and regression approaches in the explanation of syntactic variation and change in Belgian and Netherlandic Dutch. *New Ways of Analyzing Variation (NWAV 43)*. October 26, Chicago.

Grondelaers, Stefan, Roeland van Hout and Paul van Gent. 2019. Re-evaluating the Prestige of Regional Accents in Netherlandic Standard Dutch: The Role of Accent Strength and Speaker Gender. *Journal of Language and Social Psychology* 38(2). Los Angeles, CA: SAGE Publications. 215–236.

Gropen, Jess, Steven Pinker, Michelle Hollander, Richard Goldberg and Ronald Wilson. 1989. The learnability and acquisition of the dative alternation in English. *Language* 65(2). Linguistic Society of America.

Guy, Gregory. 1988. Advanced VARBRUL analysis. In K Ferrara, B Brown, K Walters & J Baugh (eds.), *Linguistic change and contact: NWAV-XVI (Texas Linguistic Forum 30)*, 124–136. University of Texas, Department of Linguistics.

Haeseryn, Walter. 2013. Belgian Dutch. In Frans Hinskens & Johan Taeldeman (eds.), *Language and space: Dutch*, 700–720. Berlin/Boston: Mouton De Gruyter.

Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij and Maarten van den Toorn. 1997. *Algemene Nederlandse Spraakkunst [General Dutch Grammar]*. Groningen: Nijhoff.

Hanks, Patrick. 2013. *Lexical analysis: norms and exploitations*. Cambridge: MIT press.

Harrell, Frank Junior. 2017. Hmisc: Harrell Miscellaneous. R package version 4.0-3.

Hawkins, John. 2002. Symmetries and asymmetries. Their grammar, typology and parsing. *Theoretical Linguistics* 28(2). 95–149.

Hawkins, John. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Heine, Bernd and Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Heine, Bernd and Tania Kuteva. 2007. *The genesis of grammar: a reconstruction*. (Studies in the Evolution of Language 9). Oxford: Oxford university press.

Heller, Benedikt. 2018. Stability and fluidity in syntactic variation world-wide. The genitive alternation across varieties of English. Dissertation University of Leuven.

Heylen, Kris, Thomas Wielfaert, Dirk Speelman and Dirk Geeraerts. 2015. Monitoring Polysemy. Word Space Models as a Tool for Large-Scale Lexical Semantic Analysis. *Lingua* 157. 153–172.

Hilpert, Martin. 2014a. *Construction Grammar and its Application to English*. Edinburgh: Edinburgh University Press.

Hilpert, Martin. 2014b. Collostructional analysis. In Dylan Glynn & Justyna Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*, 391–404. (Human Cognitive Processing [HCP]). Amsterdam: John Benjamins.

Hilpert, Martin and Susanne Flach. Forthcoming. A case of constructional contamination in English: Modified noun phrases influence adverb placement in the passive.

Hinskens, Frans & Johan Taeldeman (eds.). 2013. *Language and Space. An International Handbook of Linguistic Variation. Dutch*. Berlin, Boston: Mouton De Gruyter.

Hopper, Paul and Sandra Annear Thompson. 1980. Transitivity in Grammar and Discourse. *Language* 56(2). 251–299.

Hosmer, David and Stanley Lemeshow. 2000. *Applied logistic regression*. 2nd edn. (Wiley Series in Probability and Statistics). New York: Wiley.

Hüning, Matthias, Olivier Moliner & Ulrike Vogl (eds.). 2012. *Standard languages and multilingualism in European history*. (Multilingualism and Diversity Management 1). Amsterdam: Benjamins.

Hutchinson, Larry. 1974. Grammar as theory. *Explaining Linguistic Phenomena*, 43–73. Washington D.C.: Hemisphere Publishing Corporation.

Ibbotson, Paul, Anna Theakston, Elena Lieven and Michael Tomasello. 2012. Semantics of the Transitive Construction: Prototype Effects and Developmental Comparisons. *Cognitive Science* 36(7). Oxford, UK: Blackwell Publishing Ltd. 1268–1288.

Jaeger, Florian Tim. 2005. Optional that indicates production difficulty: Evidence from disfluencies. *Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop*, 103–109. Aix-en-Provence.

Jaeger, Florian Tim and Thomas Wasow. 2005. Production-complexity driven variation: Relativizer omission in non-subjectextracted relative clauses. *Paper presented at the 18th annual CUNY conference on sentence processing*.

Jaeger, Florian Tim, Roger Levy, Thomas Wasow and David Orr. 2005. The Absence of that is Predictable if a Relative Clause is Predictable. *Paper presented at the Architextures and Mechanisms of Language Processing conference*. Ghent.

Jaeger, Florian Tim. 2010. Redundancy and Reduction: Speakers Manage Syntactic Information Density. *Cognitive Psychology* 61(1). 23‒62.

Jaeger, Florian Tim. 2011. Corpus-based research on language production: Information density and reducible subject relatives. In Emily Bender & Jennifer Arnold (eds.), *Language from a cognitive perspective: Grammar, usage, and processing. Studies in honor of Tom Wasow*, 161‒197. Stanford: CSLI Publications.

Jaeger, Florian Tim. 2013. Production preferences cannot be understood without reference to communication. *Frontiers in Psychology* 4.

Jansen, Mathilde, Nicoline van der Sijs, Fieke der Gucht and Johan De Caluwe. 2017. *Atlas van de Nederlandse taal: Editie Nederland*. Tielt: Lannoo.

Janssens, Guy and Ann Marynissen. 2008. *Het Nederlands vroeger en nu*. 3de herw. Leuven: Acco.

Keuleers, Emmanuel and Walter Daelemans. 2007. Memory-based learning models of inflectional morphology: A methodological case study. *Lingue e Linguaggio* 6(2). 151‒174.

Kilgarriff, Adam. 1997. I don't believe in word senses. *Computers And The Humanities* 31(2). Kluwer Academic Publishers. 91‒113.

Kirby, Simon. 1999. *Function, selection, and innateness: the emergence of language universals*. Oxford: Oxford University Press.

Klavan, Jane and Dagmar Divjak. 2016. The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. *Folia Linguistica* 50(2). 355‒384.

Kloeke, Gesinus Gerhardus. 1927. *De Hollandsche expansie in de zestiende en zeventiende eeuw en haar weerspiegeling in de hedendaagsche Nederlandsche dialecten: proeve eener historisch-dialect-geographische synthese*. (Noord- En Zuid-Nederlandsche Dialectbibliotheek 2). 's-Gravenhage: Nijhoff.

Klooster, Willem Gerrit, Henk Verkuyl and Jan Luif. 1974. *Inleiding tot de syntaxis: praktische zinsleer van het Nederlands*. 5de, gewij ed. Culemborg: Stam.

Knooihuizen, Remco and Oscar Strik. 2014. Relative productivity potential of Dutch verbal inflection patterns. *Folia linguistica historica* 35(1). 173‒200.

Kollewijn, Roeland Anthonie. 1899. Voorwerpen. *Taal en Letteren* 9. 465‒480.

König, Ekkehard and Volker Gast. 2018. *Understanding English-German contrasts*. 4th edn. Berlin: Erich Schmidt.

Konopka, Agnieszka E. 2012. Planning Ahead: How Recent Experience with Structures and Words Changes the Scope of Linguistic Planning. *Journal of Memory and Language* 66(1). 143‒162.

Kraljic, Tanya and Susan Brennan. 2005. Prosodic Disambiguation of Syntactic Structures: For the Speaker or for the Addressee? *Cognitive Psychology* 50(2). 194‒231.

Krawczak, Karolina and Dylan Glynn. 2015. Operationalizing mirativity: A usage-based quantitative study of constructional construal in English. *Review of Cognitive Linguistics* 13(2). Amsterdam: John Benjamins Publishing Company.

Kruijsen, Joep and Nicoline van der Sijs. 2016. Meertens Kaartenbank. First version launched on 2014.

Labov, William. 1966a. *The social stratification of English in New York city*. Washington (D.C.): Center for applied linguistics.

Labov, William. 1966b. Hypercorrection in the lower middle class as a factor in linguistic change. In William Bright (ed.), *Sociolinguistics*, 84–113. Den Haag: Mouton.

Labov, William. 1969. Contraction, Deletion, and Inherent Variability of the English Copula. *Language: Journal of the Linguistic Society of America* 45(4). 715–762.

Labov, William. 1972a. *Sociolinguistic patterns*. (Conduct and Communication). Philadelphia: University of Pennsylvania Press.

Labov, William. 1972b. Some Principles of Linguistic Methodology. *Language in Society* 1(1). 97–120.

Lakoff, George. 1987. *Women, fire, and dangerous things. What categories reveal about the mind*. Chicago: University of Chicago Press.

Langacker, Ronald Wayne. 1990. *Concept, image, and symbol: the cognitive basis of grammar*. (Cognitive Linguistics Research 1). Berlin: de Gruyter.

Langacker, Ronald Wayne. 1991. *Foundations of cognitive grammar: descriptive application*. Stanford: Stanford University Press.

Langacker, Ronald Wayne. 2009. Constructions and constructional meaning. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New Directions in Cognitive Linguistics [Human Cognitive Porcessing]*, 225–267. Amsterdam/Philadelphia: John Benjamins.

Lavandera, Beatriz. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7(2). Cambridge University Press. 171–182.

Lehmann, Hans Martin and Gerold Schneider. 2012. Syntactic variation and lexical preference in the dative-shift alternation. In Joybrato Mukherjee & Magnus Huber (eds.), *Corpus Linguistics and Variation in English*, 65–75. Amsterdam: Rodopi.

Lemmens, Maarten. 2016. Cognitive semantics. In Nick Riemer (ed.), *Routledge Handbook of Semantics*, 90–105. London/New York: Routlegde.

Lenci, Alessandro. 2012. Argument alternations in Italian verbs: a computational study. In Valentina Bambini, Irene Ricci & Pier Marco Bertinetto (eds.), *Linguaggio e cervello - Semantica/Language and the Brain - Semantics. Atti del XLII Congresso Internazionale di Studi della Societ`a di Linguistica Italiana*, 1–26. Rome: Bulzoni.

Levin, Beth and Jason Grafmiller. 2012. Do you always fear what frightens you? In Tracy Holloway King & Valeria de Paiva (eds.), *From Quirky Case to Representing Space: Papers in Honor of Annie Zaenen*, 21–32. Stanford: CSLI Publications.

Levin, Beth. 1993. *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago press.

Levinson, Stephen. 2000. *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge: Cambridge : MIT press,.

Levshina, Natalia. 2015. *How to do linguistics with R: data exploration and statistical analysis*. Amsterdam: Benjamins.

Levshina, Natalia and Kris Heylen. 2014. A radically data-driven Construction Grammar: Experiments with Dutch causative constructions. In Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *Extending the Scope of Construction Grammar*, 17–46. Berlin: Mouton De Gruyter.

Levshina, Natalia, Dirk Geeraerts and Dirk Speelman. 2013. Towards a 3D-grammar: Interaction of linguistic and extralinguistic factors in the use of Dutch causative constructions. *Journal of Pragmatics* 52(C). 34–48.

Levy, Roger and Florian Tim Jaeger. 2007. Speakers optimize information density through syntactic reduction. In Bernhard Schölkopf, John Platt & Thomas Hoffman (eds.), *Advances in Neural Information Processing Systems 19*, 849–856. MIT Press.

Lieberman, Erez, Jean-Baptste Michel, Joe Jackson, Tina Tang and Martin Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449(7163). 713–716.

MacDonald, Maryellen and Robert Thornton. 2009. When language comprehension reflects production constraints: Resolving ambiguities with the help of past experience. *Memory & Cognition* 37(8). New York. 1177–1186.

MacDonald, Maryellen. 2013. How language production shapes language form and comprehension. *Frontiers in psychology* 4. 226.

Mars, F.K.M. 1969. Over oorzakelijke en andere voorwerpen. *De nieuwe taalgids* 62. 368–376.

Marzo, Stefania, Eline Zenner and Dorien Van de Mieroop. 2016. A prototype-theoretical approach to social meaning: s-palatalization in Cité Dutch. *Sociolinguistics Symposium*. Murcia, 15-18 June.

Marzo, Stefania. 2016. Exploring the social meaning of contemporary urban vernaculars: Perceptions and attitudes about Citétaal in Flanders. *International Journal of Bilingualism* 20(5). London, England: SAGE Publications. 501–521.

Medina, Pilar Guerrero. 2017. The English conative as a family of constructions: Towards a usage-based approach. In Francisco Ruiz de Mendoza Ibáñez, Alba Luzondo Oyón & Paula Pérez Sobrino (eds.), *Constructing families of constructions: Analytical perspectives and theoretical challenges*, 277–300. Amsterdam/Philadelphia: John Benjamins.

Menn, Lise and Cecily Jill Duffield. 2014. Looking for a 'Gold Standard' to measure language complexity: what psycholinguistics and neurolinguistics can (and cannot) offer to formal linguistics. In Frederick Newmeyer & Laurel Preston (eds.), *Measuring Grammatical Complexity*, 281–302. Oxford: Oxford University Press.

Meurers, Walt Detmar and Guido Minne. 1997. A Computational Treatment of Lexical Rules in HPSG as Covariation in Lexical Entries. *Computational Linguistics* 23(4). 543–568.

Michaelis, Laura Adrienne. 2004. Type Shifting in Construction Grammar: An Integrated Approach to Aspectual Coercion. *Cognitive Linguistics* 15(1). 1–67.

Michaelis, Laura Adrienne. 2006. Construction Grammar. In Keith Brown (ed.), *The Encyclopedia of Language and Linguistics*, 73–84. 2nd edn. v. Oxford: Elsevier.

Müller, Stefan. 2006. Phrasal or Lexical Constructions? *Language* 82(4). 850–883.

Müller, Stefan. 2018. *A lexicalist account of argument structure. Template-based phrasal LFG approaches and a lexical HPSG alternative*. Berlin: Language Science Press.

Müller, Stefan and Stephen Wechsler. 2014. Lexical approaches to argument structure. *Theoretical Linguistics* 40(1–2). 1–76.

Oehrle, Richard Thomas. 1976. The grammatical status of the English dative alternation. Cambridge: Massachusetts Institute of Technology.

Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste and Ineke Schuurman. 2013a. The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch. In Peter Spyns & Jan Odijk (eds.), *Essential Speech and Language Technology for Dutch, Theory and Applications of Natural Language Processing*, 219–247. Heidelberg: Springer.

Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste and Ineke Schuurman. 2013b. *SoNaR User Documentation*.

Oostdijk, Nelleke, Wim Goedertier, Frank Van Eynde, Louis Boves, Jean-Pierre Martens, Michael Moortgat and Harald Baayen. 2002. Experiences from the Spoken Dutch corpus project. *Proceedings of the third international conference on language resources and evaluation (LREC)*, 340–347.

Padó, Sebastian and Mirella Lapata. 2010. Dependency-Based Construction of Semantic Space Models. *Computational Linguistics* 33(2). 161–199.

Partridge, Eric. 1966. *Origins. A short etymological dictionary of modern English*. 4th ed. London: Routledge and Kegan Paul.

Pedersen, Johan. 2016. Spanish constructions of directed motion. A quantitative study. In Jiyoung Yoon & Stefan Thomas Gries (eds.), *Corpus-based approaches to Construction Grammar*, 105–144. Amsterdam/Philadelphia: John Benjamins.

Pedersen, Johan. 2019. Verb-based vs. schema-based constructions and their variability: On the Spanish transitive directed-motion construction in a contrastive perspective. *Linguistics*.

Perek, Florent. 2012. Alternation-based generalizations are stored in the mental grammar: Evidence from a sorting task experiment. *Cognitive Linguistics* 23(3). 601–635.

Perek, Florent. 2014. Rethinking constructional polysemy: The case of the English conative construction. In Dylan Glynn & Jus Robinson (eds.), *Polysemy and synonymy: Corpus methods and applications i cognitive linguistics*, 61–85. Amsterdam/Philadelphia: John Benjamins.

Perek, Florent. 2015. *Argument structure in usage-based construction grammar: experimental and corpus-based perspectives*. Amsterdam/Philadelphia: John Benjamins.

Perek, Florent. 2018. Recent change in the productivity and schematicity of the way-construction: A distributional semantic analysis. *Corpus Linguistics and Linguistic Theory* 14(1). Berlin: Walter de Gruyter GmbH. 65–97.

Perek, Florent and Maarten Lemmens. 2010. Getting at the meaning of the English at-construction: the case of a constructional split. *CogniTextes* 5. Association française de linguistique cognitive (AFLiCo).

Perek, Florent and Adele Eva Goldberg. 2015. Generalizing beyond the input: The functions of the constructions matter. *Journal of Memory and Language* 84. 108–127.

Perek, Florent and Martin Hilpert. 2017. A distributional semantic approach to the periodization of change in the productivity of constructions. *International Journal of Corpus Linguistics* 22(4). John Benjamins Publishing Company. 490–520.

Perridon, Harry. 1991. Voorzetselvoorwerpen in het Nederlands en in het Zweeds. *Tijdschrift voor Skandinavistiek* 12(1–2). 227–243.

Petré, Peter and Freek Van de Velde. 2018. The real-time dynamics of the individual and the community in grammaticalization. *Language* 94(4). 867–901.

Petré, Peter, Lynn Anthonisse, Sara Budts, William Standing Enrique Manjavacas and Odile Aurora Oscar Strik. Early Modern Multiloquent Authors (EMMA): Designing a large-scale corpus of individuals' languages. *ICAME Journal* 43.

Philippa, Marlies, Frans Debrabandere, Arend Quak, Tanneke Schoonheim and Nicoline van der Sijs. 2009. *Etymologisch woordenboek van het Nederlands*. Amsterdam: Amsterdam University Press.

Piersoul, Jozefien, Freek Van de Velde and Robbert De Troij. 150 Years of written Dutch: the construction of the Dutch corpus of contemporary and late modern periodicals (Dutch C-CLAMP).

Pijpops, Dirk. 2013. De invloed van betekenis op syntaxis bij de psychologische werkwoorden. Corpusgebaseerd alternantieonderzoek naar de syntactische variatie bij "ergeren", "interesseren", "irriteren", "storen", "verbazen", "verheugen" en "verwonderen." Master's thesis University of Leuven.

Pijpops, Dirk, Katrien Beuls and Freek Van de Velde. 2015. The rise of the verbal weak inflection in Germanic. An agent-based model. *Computational linguistics in the Netherlands Journal* 5. 81–102.

Pijpops, Dirk and Freek Van de Velde. 2016. Constructional contamination: How does it work and how do we measure it? *Folia Linguistica* 50(2). 543–581.

Pijpops, Dirk and Dirk Speelman. 2017. Alternating argument constructions of Dutch psychological verbs. A theory-driven corpus investigation. *Folia Linguistica* 51(1). 207–251.

Pijpops, Dirk and Freek Van de Velde. 2018a. Lectal contamination. How language-external variation becomes language-internal through language contact. *Variationist Linguistics meets Contact Linguisitcs*. 21 May, Ascona.

Pijpops, Dirk and Freek Van de Velde. 2018b. A multivariate analysis of the partitive genitive in Dutch. Bringing quantitative data into a theoretical discussion. *Corpus Linguistics and Linguistic Theory* 14(1). 99–131.

Pijpops, Dirk, Isabeau De Smet and Freek Van de Velde. 2018. Constructional contamination in morphology and syntax: Four case studies. *Constructions and Frames* 10(2). 269–305.

Pijpops, Dirk, Katrien Beuls and Freek de Velde. 2018. The evolutionary dynamics of language as a function of demography. *Evolang XII*. April 7, Turin.

Pijpops, Dirk, Dirk Speelman, Stefan Grondelaers and Freek Van de Velde. 2018. Comparing explanations for the Complexity Principle. Evidence from argument realization. *Language and Cognition* 10(3). 514–543.

Pinker, Steven. 1989. *Learnability and cognition : the acquisition of argument structure*. (Learning, Development, and Conceptual Change). Cambridge (Mass.): MIT press.

Plevoets, Koen. 2009. Verkavelingsvlaams als de voertaal van de verburgerlijking van Vlaanderen. *Studies van de BKL* 4. 29.

Plevoets, Koen. 2012. Van hetzelfde laken een pak: verkavelingsvlaams als uiting van status in het nieuwe rijke Vlaanderen. In Kevin Absillis, Jürgen Jaspers & Sarah Van Hoof (eds.), *De Manke Usurpator : over Verkavelingsvlaams*, 189–223. Academia Press.

Proost, Kristel. 2015. Die SUCH-Konstruktion als eine Familie von Argumentstrukturmustern. In Kristel Proost & Edeltraud Winkler (eds.), *Familienähnlichkeiten deutscher Argumentstrukturmuster*, 33–78. Mannheim: Institut für Deutsche Sprache.

Proost, Kristel. 2017. The role of verbs and verb classes in identifying German search-constructions. In Francisco Ruiz de Mendoza Ibáñez, Alba Luzondo Oyón & Paula Pérez Sobrino (eds.), *Constructing families of constructions: Analytical perspectives and theoretical challenges*, 17–52. Amsterdam/Philadelphia: John Benjamins.

Quinlan, John Ross. 1993. *C4.5: programs for machine learning*. (The Morgan Kaufmann Series in Machine Learning). San Mateo: Kaufmann.

R Core Team. 2014. *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna.

Ringe, Donald. 2006. *From Proto-Indo-European to Proto-Germanic*. Oxford: Oxford University Press.

Robinson, Justyna. 2012. A gay paper: why should sociolinguistics bother with semantics? 28(4). Cambridge University Press. 38–54.

Rohdenburg, Günter. 1996. Cognitive Complexity and Increased Grammatical Explicitness in English. *Cognitive Linguistics* 7(2). 149–182.

Rohdenburg, Günter. 2016. Testing two processing principles with respect to the extraction of elements out of complement clauses in English. *English Language and Linguistics* 20(3). 463–486.

Roland, Douglas, Jeffrey Elman and Victor Ferreira. 2006. Why is "that"? Structural Prediction and Ambiguity Resolution in a Very Large Corpus of English Sentences. *Cognition* 98(3). 245–272.

Romaine, Suzanne. 1984. On the problem of syntactic variation and pragmatic meaning in sociolinguistic theory. *Folia Linguistica* 18(3–4). 409–439.

Rosenbach, Anette. 2003. Aspects of iconicity and economy in the choice between the s-genitive and the of-genitive in English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of Grammatical Variation in English*, 379–411. Berlin/New York: Mouton De Gruyter.

Rosseel, Laura, Dirk Speelman and Dirk Geeraerts. 2018. Measuring language attitudes using the Personalized Implicit Association Test: A case study on regional varieties of Dutch in Belgium (Book review). *Journal of Linguistic Geography* 6(1). 20–39.

Röthlisberger, Melanie, Jason Grafmiller and Benedikt Szmrecsanyi. 2017. Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics* 28(4). 673–710.

Röthlisberger, Melanie. 2018a. Regional variation in probabilistic grammars: A multifactorial study of the English dative alternation. Dissertation University of Leuven.

Röthlisberger, Melanie. 2018b. Guidelines for the dative alternation.

Rueschemeyer, Shirley-Ann & Gareth Gaskell (eds.). 2018. *The Oxford Handbook of Psycholinguistics*. 2nd edn. (Oxford Handbooks Online). Oxford: Oxford University Press.

Rutten, Gijsbert, Rik Vosters and Marijke van der Wal. 2015. Frenchification in discourse and practice: Loan morphology in dutch private letters of the eighteenth and nineteenth centuries. *Past, Present and Future of a Language Border: Germanic-Romance Encounters in the Low Countries*, vol. 1, 143–169. Walter de Gruyter GmbH.

Sag, Ivan Andrew. 2012. Sign-Based Construction Grammar: an informal synopsis. In Hans Boas & Ivan Andrew Sag (eds.), *Sign-Based Construction Grammar*, 61–188. Stanford: CSLI Publications.

Sag, Ivan Andrew, Thomas Wasow and Emely Menon Bender. 2003. *Syntactic Theory: a Formal Introduction*. 2nd edn. Stanford: CSLI Publications.

Sampson, Geoffrey and Anna Babarczy. 2014. *Grammar without grammaticality. Growth and limits of grammatical precision*. Berlin: Mouton De Gruyter.

Sanders, Willy (ed.). 1971. *Expositio Willerammi Eberspergensis Abbatis in canticis canticorum: die Leidener Handschrift*. (Kleine Deutsche Prosadenkmäler Des Mittelalters 9). München: Fink.

Sankoff, Gillian. 1973. Above and beyond phonology in variable rules. In Charles-James Bailey & Roger Shuy (eds.), *New Ways of Analyzing Variation in English*, 44–62. Washington, DC: Georgetown University Press.

Scha, Renko, Rens Bod and Khalil Sima'an. 1999. A memory-based model of syntactic analysis: data-oriented parsing. *Journal Of Experimental & Theoretical Artificial Intelligence* 11(3). 409–440.

Schermer-Vermeer, Everdina. 1988. De grammatische status van het zogenaamde 'voorzetselvoorwerp'. *GLOT* 11. 11–27.

Schermer-Vermeer, Everdina. 1990. Het voorzetselvoorwerp en het begrip "vorm." *De Nieuwe Taalgids* 83. 238–247.

Schermer-Vermeer, Everdina. 1991. Een ANS zonder voorzetselvoorwerp. *Gramma* 15(3). 247–254.

Schermer-Vermeer, Everdina. 2006. Worstelen met het voorzetselvoorwerp. *Nederlandse Taalkunde* 11(2). 146–167.

Schermer-Vermeer, Everdina. 2007. Worstelen rond het voorzetselvoorwerp: een reactie op de opmerkingen van Hans Broekhuis. *Nederlandse Taalkunde* 12(4). 358–365.

Schmid, Hans-Jörg and Helmut Küchenhoff. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24(3). 531–577. (10 December, 2014).

Shank, Christopher, Julie Van Bogaert and Koen Plevoets. 2016. The Diachronic Development of Zero Complementation: A Multifactorial Analysis of the That/Zero Alternation with Think, Suppose, and Believe. *Corpus Linguistics and Linguistic Theory* 12(1). 31–72.

Shank, Christopher, Koen Plevoets and Julie Van Bogaert. 2016. A multifactorial analysis of that/zero alternation: the diachronix development of the zero complementizer with think, guess and understand. In Jiyoung Yoon & Stefan Thomas Gries (eds.), *Corpus-based Approaches to Construction Grammar*, 201–240. Amsterdam/Philadelphia: John Benjamins.

Shannon, Claude Elwoord. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27(3). Oxford, UK: Blackwell Publishing Ltd. 379–423.

Silva-Corvalán, Carmen. 1986. On the problem of meaning in sociolinguistic studies of syntactic variation. In Dieter Kastovsky & Aleksander Szwedek (eds.), *Linguistics across historical and geographical boundaries*, 111–124. Berlin/New York: Mouton De Gruyter.

Smedts, Willy and William Van Belle. 2003. *Taalboek Nederlands [Language book Dutch]*. 5th edn. Kapellen: Pelckmans.

Smith, Vicki and Herbert Clark. 1993. On the Course of Answering Questions. *Journal of Memory and Language* 32(1). 25–38.

Speelman, Dirk. 2014. Logistic regression: A confirmatory technique for comparisons in corpus linguistics. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*, 487–533. (Human Cognitive Processing [HCP]). Amsterdam: John Benjamins.

Speelman, Dirk, Stefan Grondelaers and Dirk Geeraerts. 2008. Variation in the choice of adjectives in the two main national varieties of Dutch. In Gitte Kristiansen & René Dirven (eds.), *Cognitive Sociolinguistics. Language Variation, Cultural Models, Social Systems*, 205–233. (Cognitive Linguistics Research [CLR]). Berlin/New York: Mouton De Gruyter.

Speelman, Dirk and Dirk Geeraerts. 2009. Causes for causatives: the case of Dutch "doen" and "laten." In Ted Sanders & Eve Sweetser (eds.), *Causal Categories in Discourse and Cognition*, 173–204. Berlin: Mouton De Gruyter.

Speelman, Dirk, Kris Heylen and Dirk Geeraerts. 2018. Introduction. In Dirk Speelman, Kris Heylen & Dirk Geeraerts (eds.), *Mixed-Effects Regression Models in Linguistics*, 1–10. Cham: Springer.

Speelman, Dirk, Kris Heylen and Stefan Grondelaers. A bottom-up, data-driven operationalization of semantic classes and predictability in syntactic alternation research. Forthcoming. In Stefan Grondelaers & Roeland van Hout (eds.), *New ways of analyzing syntactic variation*. Berlin: Mouton De Gruyter.

Steels, Luc. 2000. Language as a Complex Adaptive System. In Marc Schoenauer, Kalyanmoy Deb, Günter Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo & Hans-Paul Schwefel (eds.), *Proceedings of PPSN VI: Lecture notes in Computer Science*, 17–26. Berlin: Springer.

Steels, Luc. 2011a. Modeling the cultural evolution of language. *Physics of Life Reviews* 8(4). 339–356.

Steels, Luc. 2011b. *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins.

Steels, Luc. 2017. Basics of Fluid Construction Grammar. *Constructions and Frames* 9(2). 178–225.

Stefanowitsch, Anatol. 2003. Constructional semantics as a limit to grammatical alternation: The two genitives of English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of Grammatical Variation in English*, 413–441. Berlin/New York: Mouton De Gruyter.

Stefanowitsch, Anatol and Stefan Thomas Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–244.

Stefanowitsch, Anatol and Susanne Flach. 2016. The Corpus-based perspective on entrenchment. *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*, 101–127.

Stefanowitsch, Anatol. 2007. Linguistics beyond grammaticality. *Corpus Linguistics and Linguistic Theory* 3(1). Walter de Gruyter. 57–71.

Stefanowitsch, Anatol. 2010. Empirical Cognitive Semantics: some thoughts. In Dylan Glynn & Kerstin Fisher (eds.), *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, 355–380. Berlin/New York: Mouton De Gruyter.

Suffeleers, Tony. 1979. *Taalverzorging in Vlaanderen: Een opiniegeschiedenis [language care in Flanders: a history of opionions]*. (Orions Wetenschappelijke Bibliotheek). Brugge: Uitgeverij Orion.

Szmrecsanyi, Benedikt. 2005. Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1(1). Walter de Gruyter. 113–150.

Szmrecsanyi, Benedikt. 2013. The great regression: genitive variability in Late Modern English news texts. *Morphosyntactic categories and the expression of possession*, 59–88. (Linguistik Aktuell/Linguistics Today). Amsterdam: Benjamins.

Szmrecsanyi, Benedikt, Douglas Biber, Jesse Egbert and Karlien Franco. 2016. Toward more accountability: Modeling ternary genitive variation in Late Modern English. *Language Variation and Change* 28(1). 1–29.

Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller and Melanie Röthlisberger. 2016. Around the world in three alternations: modeling syntactic variation in varieties of English. *English World-Wide* 37(2). 109–137.

Szmrecsanyi, Benedikt, Jason Grafmiller, Joan Bresnan, Anette Rosenbach, Sali Tagliamonte and Simon Todd. 2017. Spoken syntax in a comparative perspective: The dative and genitive alternation in varieties of English. *Glossa* 2(1). Ubiquity Press.

Tagliamonte, Sali. 2012. *Variationist sociolinguistics: change, observation, interpretation*. (Language in Society 40). Chichester: Wiley-Blackwell.

Tanner, Darren and Nyssa Bulkes. 2015. Cues, quantification, and agreement in language comprehension. *Psychonomic Bulletin & Review* 22(6). New York. 1753–1763.

Tanner, Darren, Janet Nicol and Laurel Brehm. 2014. The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language* 76. 195–215.

Theijssen, Daphne. 2012. Making Choices. Modelling the English dative alternation. Dissertation Radboud University Nijmegen.

Thompson, Sandra Annear. 1990. Information flow and dative shift in English discourse. *Development and diversity: Language variation across time and space. A Festschrift for Charles-James N. Bailey*, 239–253. Arlington: the Summer Institute of Linguistics and the University of Texas.

Thompson, Sandra Annear and Anthony Mulac. 1991. A Quantitative Perspective on the Grammaticization of Epistemic Parentheticals in English. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Grammaticalization*, 313–339. Amsterdam: John Benjamins.

Tomasello, Michael. 2003. *Constructing a language: a usage-based theory of language acquisition*. Cambridge: Harvard university press.

Torres Cacoullos, Rena and James Walker. 2009. On the Persistence of Grammar in Discourse Formulas: A Variationist Study of That. *Linguistics: An Interdisciplinary Journal of the Language Sciences* 47(1). 1–43.

Traugott, Elizabeth Closs and Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Trost, Pavel. 1980. Präteritumsverfall und Präteritumsschwund im Deutschen. *Zeitschrift für Dialektologie und Linguistik* 47(2). Wiesbaden: Franz Steiner. 184.

Turney, Peter and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37. 141–188.

Van Belle, William and Willy Van Langendonck. 1996. The indirect object in Dutch. *The Dative. Descriptive Studies.*, 217–250. Amsterdam: John Benjamins.

Van Beveren, Amélie, Timothy Colleman and Gert De Sutter. 2019. De om-alternantie : een verkennende casestudy. *Handelingen van de Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis* 71. 191−222.

Vandekerckhove, Reinhild. 2009. Dialect loss and dialect vitality in Flanders. *International Journal of the Sociology of Language* 196−197. 73−97.

van den Bosch, Antal and Walter Daelemans. 2013. Implicit Schemata and Categories in Memory-based Language Processing. *Language and Speech* 56(3). SAGE Publications Ltd. 309−328.

van den Bosch, Antal, Stefan Grondelaers and Dirk Speelman. Forthcoming. A Memory Based Account Of Constructional Differences Between Netherlandic And Belgium Dutch. In Stefan Grondelaers & Roeland van Hout (eds.), *New Ways of Analyzing Syntactic Variation*. Berlin: Mouton De Gruyter.

van den Toorn, Maarten Cornelis. 1971. Het voorzetselvoorwerp als nominale constituent. *Studia Neerlandica* 6. 114−130.

van der Horst, Joop. 2008. *Geschiedenis van de Nederlandse syntaxis [History of Dutch syntax]*. Leuven: Universitaire Pers Leuven.

van der Horst, Joop. 2012. Aldus de mensen, incluis mij. Hoe woorden opeens voorzetsels worden. *Onze Taal* 81(10). 276−278.

van der Horst, Joop. 2013. *Taal op drift. Lange-termijnontwikkelingen in taal en samenleving [Language adrift. Long term developments in language and society]*. Amsterdam: Meulenhoff.

van der Horst, Joop and Freek Van de Velde. 2008. Het voorzetsel diachronisch. In Irena Barbara Kalla & Bożena Czarnecka (eds.), *Neerlandistische ontmoetingen. Trefpunt Wrocław*, 54−62. Oficyna Wydawnicza Atut; Wrocław.

Van de Velde, Freek. 2013. External possessors and related constructions in Functional Discourse Grammar. In John Lachlan Mackenzie & Hella Olbertz (eds.), *Casebook in Functional Discourse Grammar*, vol. 137, 155−187. Amsterdam: John Benjamins.

Van de Velde, Freek. 2014a. The Discourse Motivation for Split-Ergative Alignment in Dutch Nominalisations (and Elsewhere). *Pragmatics* 24(2). 317−348.

Van de Velde, Freek. 2014b. Degeneracy: the maintenance of constructional networks. In Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *Extending the Scope of Construction Grammar*, vol. 1, 141−179. Berlin: Mouton De Gruyter.

Van de Velde, Freek. 2017. Understanding grammar at the community level requires a diachronic perspective. Evidence from four case studies. *Nederlandse Taalkunde* 22(1). 47−74.

Van de Velde, Freek, Karlien Franco and Dirk Geeraerts. Reality check voor de kwantitatieve Nederlandse taalkunde: laveren tussen de Scylla van het conservatisme en de Charybdis van de zelfgenoegzaamheid. *Tijdschrift voor Nederlandse Taal- en Letterkunde*.

Vandeweghe, Willy and Magda Devos. 2003. Relationele subtypen bij voorzetselobjecten. *Leuvense Bijdragen* 92. 103−114.

Vandeweghe, Willy and Timothy Colleman. 2011. Drie-argumentstructuurconstructies met een voorzetselobject. *Verslagen & Mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde* 121. 205:228.

Vandeweghe, Willy. 2004. *Grammatica van de Nederlandse zin*. 3de herzie. Antwerpen: Garant.

Vandeweghe, Willy. 2011a. Het voorzetselvoorwerp en de hiërarchie der objecten. *Nederlandse Taalkunde* 16(1). 88–101.

Vandeweghe, Willy. 2011b. *Categoriën van predicaten met een voorzetselvoorwerp*.

van der Leek, Frederike. 1996. The English conative construction: A compositional account. In Lise Dobrin, Kora Singer & Lisa McNair (eds.), *Papers from the 32th Regional Meeting of the Chicago Linguistic Society*, 363–378. Chicago: Chicago Linguistic Society.

van der Wal, Marijke. 1995. *De moedertaal centraal: standaardisatie-aspecten in de Nederlanden omstreeks 1650 [The mother tongue central: standardization aspects in the Low Countries around 1650]*. (IJkpunt 1650 2). 's-Gravenhage: SDU.

van der Wal, Marijke and Cor van Bree. 2008. *Geschiedenis van het Nederlands*. Utrecht: Spectrum.

Van Eecke, Paul. 2017. Robust processing of the Dutch verb phrase. *Constructions and Frames* 9(2). 226–250.

van Hout, Anna Maria Henrica. 1996. Event semantics of verb frame alternations: a case study of Dutch and its acquisition. Dissertation Tilburg University.

van Hout, Roeland. 1989. De structuur van taalvariatie: een sociolinguïstisch onderzoek naar het stadsdialect van Nijmegen. Dissertation Radboud University Nijmegen.

Vanmassenhove, Eva, Jinhua Du and Andy Way. 2017. Investigating 'Aspect' in NMT and SMT: Translating the English Simple Past and Present Perfect. *Computational Linguistics in the Netherlands Journal* 7. 109–128.

van Noord, Gertjan, Gosse Bouma, Frank Van Eynde, Daniel De Kok, Jelmer Van der Linde, Ineke Schuurman, Erik Tjong Kim Sang and Vincent Vandeghinste. 2013. Large Scale Syntactic Annotation of Written Dutch: Lassy. In Peter Spyns & Jan Odyk (eds.), *Essential Speech and Language Technology for Dutch*, 219–247. Berlin: Springer.

van Noord, Gertjan, Ineke Schuurman and Gosse Bouma. 2011. Lassy syntactische annotatie.

van Noord, Gertjan. 2006. At Last Parsing Is Now Operational. In Piet Mertens, Cédric Fairon, Anne Dister & Patrick Watrin (eds.), *TALN 2006. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles*, 20–42. Louvain-la-Neuve: Cental.

van Santen, Ariane. 1997. Hoe sterk zijn de sterke werkwoorden. In Ariane van Santen & Marijke van der Wal (eds.), *Taal in tijd en ruimte [Feestbundel Cor van Bree]*, 45–56. Leiden.

van Trijp, Remi. 2011. A design pattern for argument structure constructions. In Luc Steels (ed.), *Design Patterns in Fluid Construction Grammar*, 115–145. Amsterdam/Philadelphia: John Benjamins.

van Trijp, Remi. 2013. A comparison between Fluid Construction Grammar and Sign-Based Construction Grammar. *Constructions and Frames* 5(1). 88–116.

van Trijp, Remi. 2014. Fitness landscapes in cultural language evolution: a case study on german definite articles. In Erica Cartmill, Seán Roberts, Headi Lyn & Hannah Cornish (eds.), *The Evolution of Language: Proceedings of the 10th International Conference*, 334–341. World Scientific.

van Trijp, Remi. 2015. Cognitive vs. generative construction grammar: The case of coercion and argument structure. *Cognitive Linguistics* 26(4). 613–632.

van Voorst, Jan. 1996. Some systematic differences between the dutch, French and English transitive construction. *Language Sciences* 18(1–2). Elsevier Ltd. 227–245.

Verhagen, Arie and Suzanne Kemmer. 1997. Interaction and causation: Causative constructions in modern standard Dutch. *Journal of Pragmatics* 27(1). Elsevier B.V. 61–82.

Verhagen, Arie. 2007. Construal and Perspectivization. *The Oxford Handbook of Cognitive Linguistics*, 48–81. Oxford: Oxford University Press.

Verhagen, Arie. 2013. Darwin en de ideale taalgebruiker. In Theo Janssen & Jan Noordegraaf (eds.), *Honderd jaar taalwetenschap. Artikelen aangeboden aan Saskia Daalder bij haar afscheid van de Vrije Universiteit.*, 151–162. Amsterdam/Münster: Stichting Neerlandistiek VU/Nodus Publikationen.

Verhagen, Véronique and Mos Maria. 2016. Stability of familiarity judgments: Individual variation and the invariant bigger picture. *Cognitive Linguistics*.

Vosters, Rik & Janneke Weijermars (eds.). 2011. *Taal, cultuurbeleid en natievorming onder Willem I.* (Verhandelingen van de Koninklijke Vlaamse Academie van België Voor Wetenschappen En Kunsten. N. R 23). Brussel: Koninklijke Vlaamse academie van België voor wetenschappen en kunsten.

Vosters, Rik and Guy Janssens. 2014. *Sur la langue nationale: taal en taalpolitiek in het Verenigd Koninkrijk der Nederlanden en het jonge België.* (Noord En Zuid Onder Willem I : 200 Jaar Verenigd Koninkrijk Der Nederlanden 4). Den Haag: Algemeen-Nederlands verbond.

Vosters, Rik. 2009. Integrationisten en particularisten? Taalstrijd in Vlaanderen tijdens het Verenigd Koninkrijk der Nederlanden. *Handelingen van de Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis* 52. 41–58.

Wallis, Sean. That vexed problem of choice. London: Survey of English Usage, UCL.

Wasow, Thomas. 2002. *Postverbal behavior.* Stanford: CSLI Publications.

Weeds, Julie, David Weir and Diana McCarthy. 2004. Characterising Measures of Lexical distributional similarity. *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*, 1015.

Weijnen, Antonius Angelus and A.P.G.M.A. Ficq-Weijnen. 2008. *Prisma woordenboek Nederlands.* 43ste druk. (Prisma Woordenboeken). Utrecht: Het Spectrum.

Weiss, Sholom and Casimir Kulikowski. 1991. *Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems.* San Mateo: Kaufmann.

Wellens, Pieter, Remi van Trijp, Katrien Beuls and Luc Steels. 2013. Fluid Construction Grammar for Historical and Evolutionary Linguistics. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 127–132. Sofia, 5 August.

Wible, David and Nai-Lung Tsao. 2017. Constructions and the problem of discovery: A case for the paradigmatic. *Corpus Linguistics and Linguistic Theory*. Published online, ahead of print.

Wickham, Hadley and Lionel Henry. 2019. tidyr: Easily Tidy Data with "spread()" and "gather()" Functions.

Wierzbicka, Anna. 1988. The semantics of "internal dative" in English. In Anna Wierzbicka (ed.), *The Semantics of Grammar*, 359–560. Amsterdam/Philadelphia: John Benjamins.

Willems, Annelore and Gert De Sutter. 2015. Reassessing the effect of the complexity principle on PP Placement in Dutch. *Nederlandse Taalkunde* 20(3). 339–366.

Wolk, Christoph, Joan Bresnan, Anette Rosenbach and Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica* 30(3). 382–419.

Zifonun, Gisela, Ludger Hoffmann and Bruno Strecker. 1997. *Grammatik der deutschen Sprache [Grammar of the German language]*. Berlin: de Gruyter.

Zipf, George Kingsley. 1949. *Human behavior and the principle of least effort : an introduction to human ecology*. New York: Hafner.

Zuidema, Willem and Arie Verhagen. 2010. What Are the Unique Design Features of Language? Formal Tools for Comparative Claims. *Adaptive Behavior* 18(1). London, England: SAGE Publications. 48–65.

Zwaan, Frederik Lodewijk. 1972. Het voorzetselvoorwerp. *Levende Talen*. 347–350.

Zwart, Jan-Wouter. 2011. *The syntax of Dutch*. Cambridge: Cambridge University Press.

# Appendix

## Chapter 6: Testing the semantic and lectal hypotheses

AIC: 23,693.0          Transitive observations: 24,346

C-index: 0.744          Prepositional observations (success level): 5,216

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | -1.84 | 0.07 | -26.32 | < 0.0001 |
| VERB | *opbellen* 'phone' (vs. *zoeken* 'search') | -5.10 | 0.87 | -5.84 | 0.0000 |
| | *schoppen* 'kick' (vs. ") | -0.87 | 0.35 | -2.51 | 0.0120 |
| | *jagen* 'hunt' (vs. ") | -0.94 | 0.71 | -1.33 | 0.1840 |
| | *bellen* 'phone' (vs. ") | -1.20 | 0.06 | -20.78 | < 0.0001 |
| | *grijpen* 'grab' (vs. ") | 0.15 | 0.07 | 2.16 | 0.0308 |
| | *vissen* 'fish' (vs. ") | 1.38 | 0.44 | 3.15 | 0.0016 |
| | *grabbelen* 'scramble' (vs. ") | 2.08 | 1.17 | 1.78 | 0.0749 |
| | *verlangen* 'desire' (vs. ") | 1.62 | 0.05 | 31.04 | < 0.0001 |
| | *telefoneer* 'phone' (vs. ") | 1.68 | 0.73 | 2.28 | 0.0225 |
| | *peilen* 'gauge' (vs. ") | -2.30 | 0.38 | -6.13 | < 0.0001 |
| | *graaien* 'grasp' (vs. ") | 3.44 | 0.50 | 6.84 | < 0.0001 |
| | *happen* 'snap' (vs. ") | 4.30 | 0.28 | 15.56 | < 0.0001 |
| THEME COMPLEXITY | | -0.12 | 0.04 | -3.30 | 0.0010 |
| VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.56 | 0.05 | 11.26 | < 0.0001 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.35 | 0.04 | 7.96 | < 0.0001 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| COMPONENT | 15 | 0.02 | 0.14 |

Table 24: Specifications of a regression model at the preposition level with VERB as fixed effect, based on the **Netherlandic** data. The categorical variables VERB and VERB-THEME ORDER are implemented through dummy coding.

AIC: 55,643.5          Transitive observations: 51,792
C-index: 0.700          Prepositional observations (success level): 12,314

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | -2.00 | 0.11 | -17.82 | < 0.0001 |
| VERB | *opbellen* 'phone' (vs. *zoeken* 'search') | -5.86 | 0.49 | -11.97 | < 0.0001 |
| | *schoppen* 'kick' (vs. ") | 0.34 | 0.27 | 1.26 | 0.2080 |
| | *jagen* 'hunt' (vs. ") | 0.40 | 0.53 | 0.75 | 0.4560 |
| | *bellen* 'phone' (vs. ") | 0.61 | 0.03 | 21.01 | < 0.0001 |
| | *grijpen* 'grab' (vs. ") | 0.94 | 0.05 | 17.81 | < 0.0001 |
| | *vissen* 'fish' (vs. ") | 1.07 | 0.22 | 4.89 | < 0.0001 |
| | *grabbelen* 'scramble' (vs. ") | 1.62 | 0.29 | 5.62 | < 0.0001 |
| | *verlangen* 'desire' (vs. ") | 1.77 | 0.06 | 31.91 | < 0.0001 |
| | *telefoneer* 'phone' (vs. ") | 2.12 | 0.14 | 15.52 | < 0.0001 |
| | *peilen* 'gauge' (vs. ") | 3.08 | 0.09 | 35.29 | < 0.0001 |
| | *graaien* 'grasp' (vs. ") | 3.15 | 0.35 | 8.94 | < 0.0001 |
| | *happen* 'snap' (vs. ") | 3.84 | 0.14 | 28.04 | < 0.0001 |
| THEME COMPLEXITY | | 0.01 | 0.02 | 0.30 | 0.7670 |
| VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.37 | 0.03 | 11.02 | < 0.0001 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.27 | 0.03 | 8.98 | < 0.0001 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| COMPONENT | 15 | 0.10 | 0.32 |

Table 25 : Specifications of a regression model at the preposition level with VERB as fixed effect, based on the **Belgian** data. The categorical variables VERB and VERB-THEME ORDER are implemented through dummy coding.

AIC: 147.44          Transitive observations: 81

C-index: 0.846       Prepositional observations (success level): 68

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | -2.35 | 0.56 | -4.17 | < 0.0001 |
| AGENT-TYPE | *helper* (vs. *aggressor*) | 3.26 | 0.54 | 6.03 | < 0.0001 |
| COUNTRY | *The Netherlands* (vs. *Belgium*) | 1.13 | 0.59 | 1.93 | 0.0537 |
| THEME COMPLEXITY | | -0.32 | 0.30 | -1.07 | 0.2863 |

Table 26: Specifications of a regression model at the object level, for *(naar) slachtoffer zoeken* 'search victim'. All categorical variables are implemented through dummy coding.

AIC: 283.16         Transitive observations: 83

C-index: 0.731       Prepositional observations (success level): 166

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | 0.62 | 0.31 | 2.00 | 0.0460 |
| WORD SPECIFICITY | *specific* (vs. *non-specific*) | -2.00 | 0.44 | -4.51 | < 0.0001 |
| COUNTRY | *The Netherlands* (vs. *Belgium*) | 0.82 | 0.30 | 2.68 | 0.0075 |
| THEME COMPLEXITY | | -0.37 | 0.18 | -2.10 | 0.0356 |
| VERB-THEME ORDER | *verb-theme* (vs. *theme-verb*) | 0.34 | 0.31 | 1.10 | 0.2695 |

Table 27: Specifications of a regression model at the object level, for *(naar) woord zoeken* 'search word'. The categorical variables are implemented through dummy coding.

# Chapter 8: Testing the complexity hypotheses

AIC: 618.3          Transitive observations: 163
C-index: 0.857      Prepositional observations (success level): 847

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | 1.62 | 0.44 | 3.68 | 0.0002 |
| THEME COMPLEXITY | | 0.00 | 0.17 | -0.02 | 0.9817 |
| VERB-THEME ORDER | *theme-verb* (vs. *verb-theme*) | -0.09 | 0.40 | -0.24 | 0.8124 |
| CLAUSE TYPE | *main clause* (vs. *sub. clause)* | 0.29 | 0.24 | 1.21 | 0.2268 |
| VERB FINITENESS | *finite* (vs. *non-finite*) | 0.94 | 0.34 | 2.79 | 0.0054 |
| | *infinitive* (vs. *participle*) | -0.64 | 0.23 | -2.75 | 0.0059 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *theme-verb* (vs. *verb-theme*) | -0.30 | 0.17 | -1.78 | 0.0744 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| THEME ROOT | 143 | 2.15 | 1.47 |
| CORPUS COMPONENT | 11 | 0.10 | 0.32 |

Table 28: Specifications of a regression model fitted on the Belgian instances of the verb *peilen* 'gauge' to test the complexity hypotheses, containing both CLAUSE TYPE and VERB FINITENESS. The categorical variables are implemented through used-defined contrast coding.

a.    theme-verb                    b. verb-theme
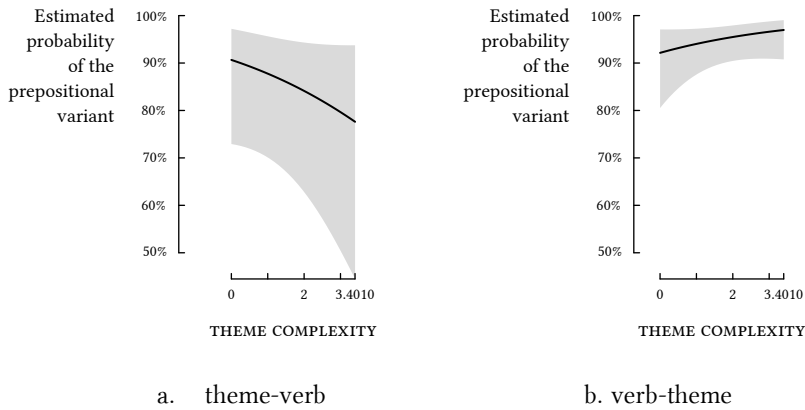
Figure 38: Effect plot of the interaction between THEME COMPLEXITY and VERB-THEME ORDER in the regression model of Belgian *peilen* 'gauge' in Table 28, containing both CLAUSE TYPE and VERB FINITENESS.

AIC: 617.7           Transitive observations: 163
C-index: 0.857        Prepositional observations (success level): 847

| Fixed effects | Level | Estimate | St. Error | Z-value | P-value |
|---|---|---|---|---|---|
| | intercept | 1.68 | 0.44 | 3.85 | 0.0001 |
| THEME COMPLEXITY | | 0.01 | 0.17 | 0.06 | 0.9536 |
| VERB-THEME ORDER | *theme-verb* (vs. *verb-theme*) | -0.34 | 0.34 | -1.00 | 0.3153 |
| VERB FINITENESS | *finite* (vs. *non-finite*) | 0.88 | 0.33 | 2.65 | 0.0080 |
| | *infinitive* (vs. *participle*) | -0.80 | 0.20 | -4.05 | < 0.0001 |
| Interaction THEME COMPLEXITY and VERB-THEME ORDER | *theme-verb* (vs. *verb-theme*) | -0.30 | 0.17 | -1.79 | 0.0743 |

| Random effects | Number of levels | Variance | Standard Deviation |
|---|---|---|---|
| THEME ROOT | 143 | 2.14 | 1.46 |
| CORPUS COMPONENT | 11 | 0.10 | 0.32 |

Table 29: Specifications of a regression model fitted on the Belgian instances of the verb *peilen* 'gauge' to test the complexity hypotheses, without CLAUSE TYPE. The categorical variables are implemented through used-defined contrast coding.
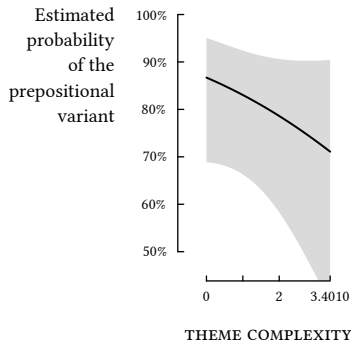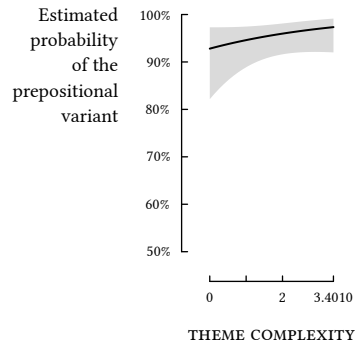
a.    theme-verb          b. verb-theme

Figure 39: Effect plot of the interaction between THEME COMPLEXITY and VERB-THEME ORDER in the regression model of Belgian *peilen* 'gauge' in Table 29, without CLAUSE TYPE.

# List of Figures

# List of Tables

# Samenvatting

Dit onderzoek heeft tot doel drie vragen te beantwoorden: hoe, waarom en waar varieert de werkwoordelijke argumentstructuur van het Nederlands met betrekking tot de alternantie tussen de transitieve en prepositoneel-intransitieve constructie. Die alternantie doet zich voor bij een aantal uiteenlopende werkwoorden en voorzetsels, zoals uit de onderstaande voorbeelden blijkt. De nadruk ligt daarbij op de werkwoorden die alterneren met het voorzetsel *naar*, zoals in (i)-(ii). De hoe-vraag gaat na welke factoren deze keuze bepalen, de waarom-vraag zoekt naar de mechanismen die de invloed van die factoren veroorzaken, en de waar-vraag bekijkt op welk niveau van abstractie deze factoren werkzaam zijn. Concreet: vinden we steeds dezelfde factoren voor alle werkwoorden, of verschillen ze van het ene werkwoord tot het andere, of zelfs van het ene object tot het andere?

i.   *We zoeken (naar) de oorzaak, maar hebben nog geen idee.*
ii.  *De initiatiefnemers peilden (naar) de behoeften van de bewoners in de wijk Sint-Gillis.*
iii. *Ze likte (aan) haar vinger en streek een wenkbrauw glad.*

Eerst wordt in kaart gebracht bij welke voorzetsels en welke werkwoorden deze alternantie zich voordoet. Daarvoor wordt gebruik gemaakt van het Sonar-corpus, dat een representatief staal bevat van het huidige geschreven Standaardnederlands in Europa. Dit levert een lijst op van 101 werkwoorden en 16 voorzetsels, oftewel 121 unieke combinaties van een werkwoord met een voorzetsel. Vervolgens zijn alle voorkomens van deze werkwoorden uit het Sonar-corpus onttrokken.

Die data zijn ingezet om een aantal hypotheses te testen uit de historische sociolinguïstiek, de cognitieve taalkunde en de psycholinguïstiek. De eerste hypotheses stellen dat de alternantie in Nederland eenvoudiger te modelleren is dan in België, en sterker bepaald wordt door betekenisverschillen en lexicale voorkeuren. De reden is dat het Belgische Nederlands een kortere periode van standaardisering gekend heeft dan het Nederlandse Nederlands. Beide voorspellingen blijken meestal te kloppen.

Om betekenisverschillen te vinden, wordt een mechanisme toegepast dat bekend staat als de lexicale-origine-hypothese uit de cognitieve taalkunde. Dit

levert een aantal concrete voorspellingen op, die getest worden aan de hand van distributionele vectoren en manuele annotatie. Vervolgens wordt Geheugen-gebaseerd Leren ingezet als data-gedreven techniek om verdere betekenisverschillen op het spoor te komen.

Ten slotte vergelijken we een aantal verklaringen voor de invloed van complexiteit op taalvariatie. Een eerste verklaring zoekt de fundamentele oorzaak daarvoor in de cognitieve taalverwerking van de taalproducent, een tweede in de fysieke beperkingen van het taalkanaal, en een derde in de cognitieve taalverwerking van de taalontvanger. De voorspellingen van de eerste en tweede verklaring worden bevestigd, die van de derde weerlegd.

Deze invloed van taalcomplexiteit, alsook die van het onderscheid tussen België en Nederland, blijkt vrij stabiel op verschillende niveaus van abstractie. In Nederland vinden we zowel sterkere lexicale voorkeuren voor een variant per werkwoord en per object. De semantische factoren blijken echter voornamelijk op de lagere niveaus te spelen. We stellen verscheidene betekenisverschillen vast, die variëren van de ene groep werkwoorden tot de andere groep, van het ene werkwoord tot het andere, en soms zelfs van het ene object tot het andere. Het ontstaan van deze betekenisverschillen kan echter wel verklaard worden via dezelfde mechanismen, zoals de lexicale-origine-hypothese.

Deze thesis laat zien dat variatie in argumentstructuur op verscheidene niveaus van abstractie bepaald wordt, dat de betekenisverschillen die veroorzaakt worden door het voorzetsel *naar* weliswaar uiteenlopend zijn, maar zeker niet onvoorspelbaar, en dat het bij variatieonderzoek cruciaal is inzichten te combineren uit verschillende takken van de taalkunde.

# Abstract

This investigation aims to answer three questions: how, why and where does verbal argument structure vary with respect to the alternation between the transitive and prepositional-intransitive construction in Dutch? This alternation crops up with a number of disparate verbs and prepositions, as may be apparent from the examples below. The study focusses on alternations with the preposition *naar* 'to', as in (i)-(ii). The how-question asks which factors determine the choice between both variants, the why-question pertains to the mechanisms that cause the influence of these factors, and the where-question inquires at what level of abstraction these factors operate. Put concretely, do we find the same factors for all verbs, or do they differ from one verb to the next or even from one object to the next?

i.  *We zoeken (naar) de oorzaak, maar hebben nog geen idee.*
    we search (to) the cause but have yet no idea
    'We are searching for the cause, but do not yet have a clue.'

ii.  *De initiatiefnemers peilden (naar) de behoeften van de bewoners*
    the initiators gauged (to) the needs of the inhabitants
    *in de wijk Sint-Gillis.*
    in the neighborhood Sint-Gilles
    'The initiators gauged the needs of the inhabitants in the neighborhood Sint-Gilles.'

iii.  *Ze likte (aan) haar vinger en streek een wenkbrauw glad.*
    she licked (on) her finger and brushed a eyebrow smooth
    'She licked her finger and smoothened an eyebroy.'

First, we map out which prepositions and which verbs exhibit the alternation. For this, we make use of the Sonar-corpus, which contains a representative sample of contemporary written Standard Dutch in Europe. This yields a list of 101 verbs and 16 prepositions, or 121 unique combinations of a verb and a prepositions. Next, all instances of these verbs are extracted from the Sonar-corpus.

The data are then used to test a number of hypotheses from historical sociolinguistics, cognitive linguistics and psycholinguistics. The first hypotheses predict that the alternation should be more easy to model in the Netherlands than in Belgium, and would be more strictly determined by meaning differences and lexical preferences. The reason would be that Belgian Dutch went through a shorter period of standardization than Netherlandic Dutch. These predictions are generally confirmed.

To find meaning differences, we apply a mechanism known as the lexical origin hypothesis from cognitive linguistics. A number of concrete predictions are derived from this, which are tested via the use of distributional vectors and manual annotation. Next, Memory-based Learning is employed as a data-driven technique to track down any further meaning differences.

Finally, we compare a number of explanations for the influence of complexity on language variation. According to a first explanation, the fundamental cause of this influence should be sought in the cognitive processing done by the language producer, a second points to the physical restrictions of the language channel, while a third proposes to look for it in cognitive comprehension processing. The predictions of the first and the second explanation are confirmed, those of the third are refuted.

This influence of language complexity, as well as the distinction between Belgium and the Netherlands, appears fairly stable across various levels of abstraction. In the Netherlands, we find both stronger lexical biases for one of the variants per verb and per object. Conversely, the semantic factors seems mostly at play at the lower levels. We observe various meaning differences, that vary from one group of verbs to the next from one verb to the next, and sometimes even from one object to the next. Still, the emergence of these meaning differences can be explained through the same mechanisms, such as the lexical origin hypothesis.

This dissertation shows (i) that variation in argument structure is determined at various levels of abstraction, (ii) that meaning differences that are caused by the presence of the preposition *naar* 'to' are diverse, to be sure, but certainly not unpredictable, and (iii) how it is crucial to combine insights from several disciplines to study language variation.