

Frank Van Eynde*

Clustering and stranding in Dutch

<https://doi.org/10.1515/ling-2019-0023>

Abstract: This paper has both a descriptive and a theoretical aim. The descriptive one is to demonstrate that the phenomenon of clustering is not limited to verbs, but that it also affects adpositions. The theoretical one is to develop a formal analysis that captures the common properties of verb clustering and adposition clustering. For that purpose we employ the framework of Head-driven Phrase Structure Grammar. Both the descriptive and the theoretical part are backed up with quantitative data about the use of adposition clusters in a sample consisting of one million words of spoken Dutch and one million words of written Dutch.

Keywords: circumposition, particle, adposition stranding, adposition cluster, verb cluster, complement raising, complement extraction, Head-driven Phrase Structure Grammar

1 Introduction

Van Eynde (2017) argues that Dutch not only allows the formation of verb clusters but also of adposition clusters. The latter take the form of stranded circumpositions and have a number of properties in common with verb clusters. This paper provides an update of that descriptive study (Sections 2 and 3) and demonstrates that the phenomena of clustering and stranding are interrelated, in the sense that they result from the same syntactic processes, i.e. complement raising and complement extraction.¹ The demonstration takes the form of an analysis which is cast in the notation of Head-driven Phrase Structure Grammar (Section 4). The analysis is compared with other monostratal treatments of clustering and stranding (Section 5), and the main findings are summarized in the conclusion (Section 6).

1 In the terminology of transformational grammar complement raising is a form of scrambling and complement extraction a form of A-bar-movement.

***Corresponding author: Frank Van Eynde**, Department of Linguistics, KU Leuven, Blijde-Inkomststraat 21 box 3315, 3000 Leuven, Belgium, E-mail: frank.vaneynde@kuleuven.be

Both the descriptive and the theoretical part are backed up with quantitative data about the use of verb clusters and adposition clusters in a sample consisting of one million words of spoken Dutch and one million words of written Dutch.²

2 Circumpositions

Dutch adpositions may precede their complement, as in (1), or follow it, as in (2).

- (1) *In augustus 1864 begon Sheridan met een strooptocht door de vallei*
 P August 1864 began Sheridan P a robbing.raid P the valley
 valley

‘In August 1864 Sheridan started a raid through the valley’
 (wr-p-e-i-0000050394.p.9.s.154)

- (2) *De hele dag door komen er mensen om het register te tekenen*
 the whole day P come there people the register to sign
 sign

‘All day long people are coming in order to sign the register’
 (ws-u-e-a-0000000014.p.9.s.1)

Adopting standard practice, we call them PREpositions and POSTpositions respectively. A third possibility for the Dutch adpositions is to surround their complement, as in (3).

- (3) *Tegen de eerdere afspraken in krijgt Milosevic ... spreektijd*
 P₁ the earlier agreements P₂ gets Milosevic ... speech.time
 ‘Against the earlier agreements Milosevic gets speaking time’
 (ws-u-e-a-0000000241.p.38.s.9)

We call them CIRCUMpositions. Some other examples are *door ... heen* and *achter ... aan*, as used in (4–5).

² For the spoken data we use the CGN treebank (version 2.0.1), described in Oostdijk et al. (2002). For the written data we use LASSY small (version 1.1), described in Van Noord et al. (2013). Examples which are extracted from these treebanks have a unique identifier. Those from the CGN treebank have the ‘fn’ prefix if they are from the Netherlands and the ‘fv’ prefix if they are from the Dutch speaking part of Belgium (Vlaanderen).

- (4) *Door de jaren heen is niet goed met die verdragsgelden*
 P₁ the years P₂ is not good P those treaty.moneys
omgesprongen
 handled
 ‘Throughout the years those treaty funds have not been handled well’
 (wr-p-p-h-0000000023.p.6.s.1)
- (5) *Die zitten achter die kippen aan*
 they sit P₁ the chickens P₂
 ‘They are after the chickens’
 (fni007258_38)

This section provides an inventory of the Dutch circumpositions (2.1), a set of criteria to differentiate them from separable verb particles and predicative adpositions (2.2), and an analysis of the internal structure of circumpositional PPS (2.3).

2.1 An inventory of circumpositions

Inventories of circumpositions can be found in descriptive grammars, such as Haeseryn et al. (1997: 526–530) and Broekhuis (2013: 50–51). The former lists 32 circumpositions, the latter 36, each exemplified by self-made sentences. To collect some data about usage and frequency we did a search for circumpositions in the two million word sample. This is facilitated by the fact that the annotation manual prescribes the use of a separate dependency label for the second part of a circumposition (@rel=“hdf”) (Hoekstra et al. 2003). Making use of GrETEL 2.0, a tool for example-based treebank querying (Augustinus et al. 2012), we derived XPath queries which retrieve all PPS which consist of an adposition, an object complement and another adposition, in that order.³ They yield 270 hits for LASSY small and 571 for the CGN treebank. Table 1 provides a survey of the relevant hits.⁴ They are classified according to the rightmost adposition. This yields 11 classes. Most frequent are the combinations with *toe* (333 hits) and *heen* (183 hits). Some of the circumpositions consist of more than two adpositions, as in (6).

³ The format of the queries is spelled out in the Appendix.

⁴ The sum of the relevant hits is 774, which is 67 less than the total number of hits (270 + 571). The difference is explained in the Appendix.

Table 1: The Dutch circumpositions in the sample.

<i>achter ... aan</i>	13	<i>door ... heen</i>	47	<i>naar ... toe</i>	166
<i>op ... aan</i>	1	<i>langs ... heen</i>	4	<i>op ... toe</i>	2
<i>tegen ... aan</i>	28	naar ... heen	4	<i>tot ... toe</i>	165
<i>bij ... af</i>	1	<i>om ... heen</i>	69	<i>boven ... uit</i>	7
<i>op ... af</i>	5	<i>over ... heen</i>	59	door ... uit	1
<i>van ... af</i>	51	bij ... in	2	naar ... uit	2
aan ... door	3	<i>tegen ... in</i>	17	<i>onder ... uit</i>	1
achter ... door	2	<i>tussen ... in</i>	15	<i>van ... uit</i>	34
<i>onder ... door</i>	8	<i>achter ... langs</i>	3	<i>voor ... uit</i>	10
<i>tussen ... door</i>	13	<i>buiten ... om</i>	7	<i>bij ... vandaan</i>	6
		<i>tegen ... op</i>	6	<i>onder ... vandaan</i>	8
		van ... op	1	<i>van ... vandaan</i>	3
<i>van ... afaan</i>	5	<i>tot aan ... toe</i>	1	<i>van onder ... vandaan</i>	1
		<i>tot in ... toe</i>	3		

- (6) *Hoorn was bijna van 't begin af aan bij de*
 Hoorn was nearly P_1 the beginning P_2 P_3 P the
walvisvaart betrokken
 whale.fishing involved
 'Hoorn was involved in whale hunting nearly from the beginning'
 (fnj007418_100)

They are listed separately below the double line. The combinations in bold do not figure in either Haeseryn et al. (1997: 526–530) or Broekhuis (2013: 50–51). A common property of these circumpositions is that they exclusively occur in the Spoken Dutch Corpus, mostly in the components which consist of spontaneous face-to-face conversations (fna and fva). An example is *naar ... heen*, as used in (7).

- (7) *als je dus naar Mexico heen gaat*
 if you thus P_1 Mexico P_2 go
 'if you are going to Mexico'
 (fna000570_37)

The absence from Haeseryn et al. (1997: 526–530) and Broekhuis (2013: 50–51) is probably due to the fact that they focus on Standard Dutch.⁵

Conversely, some of the circumpositions that are listed in Haeseryn et al. (1997: 526–530) and Broekhuis (2013: 50–51) do not occur in the sample. Several of

⁵ In the case of *aan ... door* it may also be due to the fact that it only occurs in the multi-word unit *aan één stuk door* 'uninterruptedly'.

them concern combinations with *langs* and *om*, which are the least common in the sample with respectively 3 and 7 hits (*boven/onder/voor ... langs* and *achter/voor ... om*). Also missing is a number of combinations with *uit* (*achter/tussen ... uit*) and *vandaan* (*achter/om/tussen/uit/voor ... vandaan*). Their absence is probably due to data sparseness.

2.2 Identifying circumpositional PPs

Not every [P–NP–P] sequence is a circumpositional PP. To illustrate this, let us compare the [*van*–NP–*af*] sequences in (8–9).

- (8) *Van het begin af hadden de Japanners verwacht dat ...*
 P₁ the beginning P₂ had the Japanese expected that ...
 ‘The Japanese had expected right from the start that ...’
 (ws-u-e-a-0000000020.p.44.s.18)

- (9) ... *zag het Huis Wittelsbach van zijn aanspraken op de*
 ... saw the House Wittelsbach P its pretenses P the
keizerstroom af
 emperor.throne SVP
 ‘...the House Wittelsbach gave up its claims to the throne of the emperor’
 (wr-p-e-i-0000054957.p.2.s.110)

In (8) *af* is part of the circumpositional PP *van het begin af*, but in (9) it is the separable particle of the verb *afzien* ‘give up’, which takes a PP-complement that is introduced by *van*. To differentiate them we use three diagnostic tests. The first concerns topicalization: While a circumpositional PP can be topicalized, as in (3–4) and (8), a sequence of a PP and a separable verb particle cannot, as shown in (10).

- (10) * *van zijn aanspraken op de keizerstroom af zag het*
 * P its pretenses P the emperor.throne SVP saw the
Huis Wittelsbach
 House Wittelsbach

This suggests that a circumpositional PP is a constituent while a sequence of a PP and a separable verb particle is not. Confirming evidence is provided by a second test: While a circumpositional PP can be used adnominally, as in (11), a sequence of a PP and a separable verb particle cannot.

- (11) *de verspreiding van deze foto door de jaren heen heeft ...*
 the distribution P this picture P₁ the years P₂ has ...
 ‘The distribution of this picture throughout the years has ...’
 (wr-p-e-i-0000004745.p.5.s.101)
- (12) * *het zien van haar aanspraken af heeft ons erg verbaasd*
 * the seeing P her pretenses SVP has us much surprised

A third test is based on the linear order of the adpositions. In a circumpositional PP that order is fixed. It is not possible to put P₂ before P₁, as shown by the contrast in (13).

- (13) *De Japanners hadden van het begin af verwacht dat ...*
 the Japanese had P₁ the beginning P₂ expected that ...
 ‘The Japanese had expected right from the start that ...’
- (14) * *De Japanners hadden af verwacht van het begin dat ...*
 * the Japanese had P₂ expected P₁ the beginning that ...

In a sequence of a PP and a separable verb particle, by contrast, the order is not fixed. The particle may follow the PP, as in (9), but it may also precede it, as in (15).

- (15) ... *zag het Huis Wittelsbach af van zijn aanspraken op de*
 ... saw the House Wittelsbach SVP P its pretenses P the
keizerstroom
 emperor.throne
 ‘...the House Wittelsbach gave up its claims to the throne of the emperor’

The first two tests are also used in Broekhuis (2013: 52–66) (his numbers IV and V) and the the third one is a generalization of his PP-over-V test (III). The two remaining tests in Broekhuis (2013) are based on the omissibility of the [P–NP] sequence (I) and on its substitutability by a locative pronoun (II), but the application of these tests is accompanied by so many ifs and buts that we leave them aside.

More relevant at this point is the fact that the three diagnostic tests can also be used to differentiate circumpositional PPs from other sequences of a PP and an adposition that do not form a constituent. A relevant case is (16), where *af* is the predicative complement of *moet* ‘must’.⁶

⁶ In contrast to the English modals, the Dutch modals are compatible with a predicative complement. See Van Eynde (2015: 213–214) for examples and discussion.

- (16) *die moet weliswaar van haar afval af*
 that must indeed P her rubbish P
 ‘she must indeed get rid of her rubbish’
 (ws-u-e-a-0000000015.p.31.s.6)

That this is not an instance of the circumposition *van ... af* is clear from the fact that the [*van-NP-af*] sequence cannot be topicalized nor used in adnominal position.

- (17) * *van haar afval af moet die weliswaar*
 * P its rubbish P must that indeed
- (18) * *de verplichting van haar afval af heeft de firma veel*
 * the obligation P her rubbish P has the firm much
geld gekost
 money cost

Confirming evidence is provided by the fact that *af* may precede the PP complement of the verb, as in (19).

- (19) *de twee ... politici wilden ook al af van die*
 the two ... politicians wanted also already P P that
communautaire onzin
 community nonsense
 ‘Also the two politicians wanted to get rid of that community nonsense’
 (wr-p-p-i-0000000263.p.4.s.5)

Summing up, circumpositional PPs are constituents that can be topicalized and used adnominally, and the linear order of the adpositions in a circumposition is fixed. This differentiates them from sequences of a PP and a separable verb particle or a predicative adposition.

2.3 The internal structure of circumpositional PPs

Within a circumpositional PP the NP complement is more tightly related to the first adposition than to the second one. This is clear from the fact that the [P_1 -NP] sequence can be separated from P_2 , as in (20), whereas the [NP - P_2] sequence cannot be separated from P_1 , as shown in (21).⁷

⁷ Notice that (21) violates the constraint that P_2 must follow P_1 .

(20) *achter welke optocht liepen de kinderen __ aan?*

P₁ which parade ran the children __ P₂?

‘which parade did the children run after?’

(21) * *welke optocht aan liepen de kinderen achter __?*

* which parade P₂ ran the children P₁ __?

Some instances of P₂ stranding from the sample are (22–23).⁸

(22) *dat 3000 boeken ... niet door het conversieproces waren __*

that 3000 books ... not P₁ the conversion.process were __

heen gekomen

P₂ come

‘that 3000 books had not made it through the conversion process’

(wr-p-p-e-000000002.p.18.s.3)

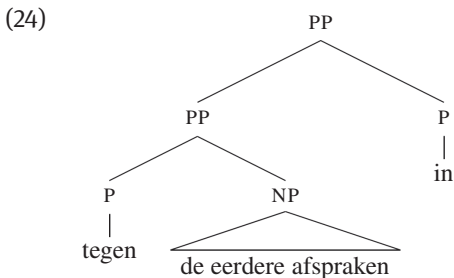
(23) *achter Miranda loop ik dan __ langs*

P₁ Miranda run I then __ P₂

‘I then run around behind Miranda’

(fna000817_39)

To capture this we assume that circumpositional PPs have a left branching structure, as in (24).



Further evidence for this structure is provided by the fact the sample contains several instances in which P₂ scopes over a sequence of [P₁–NP] conjuncts, as in (25), but none in which P₁ scopes over a sequence of [NP–P₂] conjuncts.

⁸ The rightmost adposition in (22) is stranded in the verb cluster. For a treatment of cluster creepers with quantitative data from the sample, see Augustinus and Van Eynde (2014).

- (25) *we hebben ... tegen de klok en tegen mekaar in zitten*
 we have ... P₁ the clock and P₁ each.other P₂ sit
lezen
 read
 ‘we have been reading against the clock and each other’
 (fna000451_181)

The binary structure is also applicable to circumpositions with more than two members, as in (26) and (6), repeated in (27).

- (26) ... *worden alle verdwijningsdossiers van onder het stof vandaan*
 ... become all disappearance.files P₁ P₂ the dust P₃
gehaald
 got
 ‘all disappearance files are taken out from under the dust’
 (wiki-356.p.7.s.2)

- (27) *Hoorn was bijna van 't begin af aan bij de*
 Hoorn was nearly P₁ the beginning P₂ P₃ P the
walvisvaart betrokken
 whale.fishing involved
 ‘Hoorn was involved in whale hunting nearly from the beginning’

The relevant structures are given in (28).

- (28)
-

The structure at the left applies to combinations with two prepositions and one postposition. It is not only attested by *van onder ... vandaan* (1 hit), but also by *tot aan ... toe* (1 hit) and *tot in ... toe* (3 hits). Evidence for treating the lowest PP as a complement of the first adposition is provided by the fact that *van* and *tot* belong to the small class of PREpositions which can take a PP complement, as in

van voor de oorlog ‘from before the war’ and *tot aan de grens* ‘till at the border’.⁹ The structure at the right applies to combinations with one preposition and two postpositions. Evidence for treating *van ... af* as a constituent is provided by the fact that it also occurs without *aan*, as in (29).

- (29) ... *die van het begin af de terreur-versie afwees*
 ... which P₁ the beginning P₂ the terror-version rejected
 ‘... which rejected the terror version from the beginning’
 (ws-u-e-a-000000237.p.16.s.2)

Taking stock, circumpositional PPs have a binary branching structure in which the rightmost adposition heads the higher PP and takes the lower PP as its complement. As usual in such combinations, the selecting adposition constrains the choice of the adposition in the selected PP. *Af*, for instance, selects a PP which is introduced by *bij*, *op* or *van*.

3 R-pronouns and adposition stranding

A peculiar property of the Dutch adpositions is that they do not combine with a number of singular neuter pronouns. The demonstrative *dat* ‘that’, for instance, cannot be used as the complement of an adposition, as in (30). In that combination it must be replaced by the R-form *daar* ‘that.R’. Moreover, it must precede the adposition, as in (31).

- (30) * *dat ze nog vaak aan dat denkt*
 * that she still often P that thinks

- (31) *dat ze nog vaak daar aan denkt*
 that she still often that.R P thinks
 ‘that she often thinks of that’

The same holds for *dit* ‘this’, *het* ‘it’ and *wat* ‘what’.¹⁰ For the quantifying pronouns *iets* ‘something’, *niets* ‘nothing’ and *alles* ‘everything’, the replacement is optional: (32) and (33) are equally well-formed.

⁹ The third member of this small class is *voor* ‘for’, as in *voor bij de koffie* ‘for with the coffee’.

¹⁰ The R-form pronouns have to be distinguished from the homophonous locative adverbs. In *dat ze niet van hier zijn* ‘that they are not from here’, *hier* is not the R-form of the pronoun *dit*, but a locative adverb, meaning ‘here’. As the example shows, the locative adverbs follow the adposition, just like ordinary NPs.

- (32) *dat ze soms gewoon aan niets denkt*
 that she sometimes simply P nothing thinks
 ‘that she simply thinks of nothing from time to time’
- (33) *dat ze soms gewoon nergens aan denkt*
 that she sometimes simply nothing.R P thinks
 ‘that she simply thinks of nothing from time to time’

Table 2 provides a survey of the pronouns which show the [-/+ R] alternation.

Table 2: The Dutch pronouns with an R-form.

	[-R]	[+R]		[-R]	[+R]
that	<i>dat</i>	<i>daar</i>	everything	<i>alles</i>	<i>overal</i>
this	<i>dit</i>	<i>hier</i>	something	<i>iets</i>	<i>ergens</i>
it	<i>het</i>	<i>er, d'r</i>	nothing	<i>niets</i>	<i>nergens</i>
what	<i>wat</i>	<i>waar</i>			

The alternation has no effect on the form of the adposition in (30–33), but for some of the adpositions there is an effect. *Met* ‘with’ and *tot* ‘till’, for instance, invariably precede their complement and are, hence, not compatible with an R-pronoun. In that combination they are replaced by respectively *mee* and *toe*, as in *daar mee/*met* and *er toe/*tot*.

Characteristic of the R-pronouns is that they may be realized out of the local PP, leaving the adposition stranded, as in (34).

- (34) *dat ze daar nog vaak ___ aan denkt*
 that she that.R still often ___ P thinks
 ‘that she often thinks of that’

This phenomenon has been studied extensively. Descriptive surveys are provided in Haeseryn et al. (1997) and Broekhuis (2013), transformational treatments in Van Riemsdijk (1978) and Bennis (1986), and monostratal treatments in Rentier (1993), Bouma (2000) and Van Eynde and Augustinus (2014).

3.1 Stranded circumpositions

Of special interest in this paper is the use of R-pronouns in combination with a circumposition. This use is unexceptional: In the combination of *het* ‘it’ with *naar ... toe*, for instance, the pronoun takes the R-form and precedes its selector, as in (35).

- (35) ... *op eigen gelegenheid er naar toe komen*
 ... P own opportunity it.R P₁ P₂ come
 ‘... come to it by one’s own means’
 (wr-p-e-i-0000049645.p.1.s.156)

Moreover, it may be realized out of the local PP, as in (36).

- (36) *toch trekken borstkankerspecialisten er jaarlijks — naar toe*
 yet go breast.cancer.specialists it.R yearly — P₁ P₂
 ‘still breast cancer specialists go there every year’
 (dpc-med-000684-nl-sen.p.4.s.1)

The sequence *naar toe* in (36) is, hence, a stranded circumposition.

To find out which of the circumpositions can be stranded and to collect some data about frequency and usage we turned again to the two million word sample. Unfortunately, retrieving the relevant instances was less straightforward than anticipated, for while the distinctive dependency label for the second part of a circumposition (@rel=“hdf”) is assigned with considerable accuracy if the circumposition surrounds its complement, it is not assigned consistently if the circumposition follows its complement, especially if it is separated from it, as in (36). For that reason we ignored the annotation and resorted to string search by regular expressions. A second complication is the inconsistency of the orthography: While the two parts of a stranded circumposition are canonically treated as separate words, as in (35–36), there is a tendency to amalgamate them when they are adjacent, as in (37).

- (37) *mocht u er nog naartoe willen*
 might you it.R still P₁.P₂ want
 ‘in case you might want to go there.’
 (ws-u-e-a-0000000005.p.37.s.30)

In spite of what the spelling suggests, the syntactic and semantic relations between *er*, *naar* and *toe* in (37) are the same as in (35). Making matters worse, the inconsistency of the orthography also affects the R-pronouns: While they are canonically treated as separate words, there is tendency to incorporate them in the adposition when they are adjacent, as in (38–39).

- (38) *welk publiek komt daarop af?*
 which audience comes that.R.P₁ P₂
 ‘what kind of audience comes to that?’
 (fna000254_10)

- (39) *de mensen die verplicht werden om ernaartoe te gaan*
 the people who obliged were it.R.P₁.P₂ to go
 ‘the people who were obliged to go there’
 (wr-p-e-i-0000050381.p.1.s.568)

Also here, there is no difference in meaning or syntactic function between the instances with an incorporated pronoun and those in which the pronoun is treated as a separate orthographic unit. For that reason each query had to be applied in four variants. Doing that for the combinations in Table 1 as well as for the twelve non-attested ones (listed in the last paragraph of Section 2.1), we got more than 1700 hits. Given the crude nature of string search it is not surprising that many of these are false hits. They were filtered out manually.¹¹ Table 3

Table 3: Adposition clusters in the sample.

canonical	40	amalgam	286	R-incorporation	8	amalgam and R-incorporation	98	432
<i>achter aan</i>	0	<i>achteraan</i>	12	<i>Rachter aan</i>	0	<i>Rachteraan</i>	4	16
<i>op aan</i>	2	<i>opaan</i>	0	<i>Rop aan</i>	0	<i>Ropaan</i>	0	2
<i>tegen aan</i>	0	<i>tegenaan</i>	19	<i>Rtegen aan</i>	0	<i>Rtegenaan</i>	8	27
<i>op af</i>	7	<i>opaf</i>	0	<i>Rop af</i>	2	<i>Ropaf</i>	0	9
<i>van af</i>	10	<i>vanaf</i>	10	<i>Rvan af</i>	4	<i>Rvanaf</i>	0	24
<i>onder door</i>	0	<i>onderdoor</i>	3	<i>Ronder door</i>	0	<i>Ronderdoor</i>	4	7
<i>tussen door</i>	0	<i>tussendoor</i>	8	<i>Rtussen door</i>	0	<i>Rtussendoor</i>	4	12
<i>door heen</i>	0	<i>doorheen</i>	42	<i>Rdoor heen</i>	0	<i>Rdoorheen</i>	6	48
<i>om heen</i>	2	<i>omheen</i>	13	<i>Rom heen</i>	0	<i>Romheen</i>	5	20
<i>over heen</i>	2	<i>overheen</i>	26	<i>Rover heen</i>	1	<i>Roverheen</i>	3	32
<i>bij in</i>	3	<i>bijin</i>	0	<i>Rbij in</i>	0	<i>Rbijin</i>	0	3
<i>tegen in</i>	0	<i>tegenin</i>	2	<i>Rtegen in</i>	0	<i>Rtegenin</i>	1	3
<i>tussen in</i>	0	<i>tussenin</i>	4	<i>Rtussen in</i>	0	<i>Rtussenin</i>	4	8
<i>achter langs</i>	1	<i>achterlangs</i>	0	<i>Rachter langs</i>	0	<i>Rachterlangs</i>	0	1
<i>boven langs</i>	1	<i>bovenlangs</i>	0	<i>Rboven langs</i>	0	<i>Rbovenlangs</i>	0	1
<i>tegen op</i>	0	<i>tegenop</i>	1	<i>Rtegen op</i>	0	<i>Rtegenop</i>	0	1
<i>naar toe</i>	10	<i>naartoe</i>	136	<i>Rnaar toe</i>	0	<i>Rnaartoe</i>	56	202
<i>op toe</i>	1	<i>optoe</i>	0	<i>Rop toe</i>	0	<i>Roptoe</i>	0	1
<i>achter uit</i>	0	<i>achteruit</i>	1	<i>Rachter uit</i>	0	<i>Rachteruit</i>	0	1
<i>onder uit</i>	1	<i>onderuit</i>	3	<i>Ronder uit</i>	1	<i>Ronderuit</i>	1	6
<i>tussen uit</i>	0	<i>tussenuit</i>	5	<i>Rtussen uit</i>	0	<i>Rtussenuit</i>	2	7
<i>van uit</i>	0	<i>vanuit</i>	1	<i>Rvan uit</i>	0	<i>Rvanuit</i>	0	1

¹¹ Details about the retrieval and the filtering are given in the Appendix.

provides a survey of the relevant hits, sorted on the basis of the rightmost adposition. We find the same 11 groups as in Table 1, except for the fact that two are missing, i.e. those with *om* and *vandaan*. Examples can be constructed, though, such as (40), or found in larger corpora, such as (41), which is extracted from the SoNaR corpus (Oostdijk et al. 2013).

(40) *we konden er nog net __ buiten om*
 we could it.R still just __ P₁ P₂
 ‘we just about managed to get around it’

(41) *hij kwam er een paar keer __ achter vandaan*
 he came it.R a pair time __ P₁ P₂
 ‘he came from behind it a couple of times’

The sample does not contain any combinations with more than two adpositions either, but they can be constructed as well.¹²

(42) *waar een vrouw levend van __ onder uit gehaald werd*
 what.R a woman alive P₁ __ P₂ P₃ taken was
 ‘from under which a woman was taken alive’

(43) *waar hebben ze die dossiers van __ onder vandaan*
 what.R have they those files P₁ __ P₂ P₃
gehaald?
 taken?
 ‘what did they take those files from under?’

The only circumpositions for which the stranding is impossible are those which are introduced by an adposition that is invariably PREpositional. They include *tot toe* and its ternary variants *tot aan toe* and *tot in toe*. They cannot be stranded, since *tot* cannot be preceded by its complement.

3.2 Clustering

At this point we are ready to describe the similarity between stranded circumpositions and verb clusters. An example of the latter is given in (44).

(44) *dat we die trein beter niet __ genomen hadden*
 that we that train better not __ taken had
 ‘that we had better not taken that train’

¹² (42) is quoted from Smessaert et al. (2014: 164).

The auxiliary in this clause takes a participial VP as its complement and that complement is headed by *genomen* ‘taken’, but the latter is separated from its object complement *die trein* ‘that train’ by the adjunct *beter niet* ‘better not’. As a result, the auxiliary and the participle form a cluster that is separated from the non-verbal complements of the verbs.

This mirrors the relations in (45).

- (45) *kan ze overal op de fiets ___ naar toe*
 can she all.R P the bike ___ P₁ P₂
 ‘she can get everywhere by bike.’
 (ws-u-e-a-0000000222.p.20.s.1)

The postpositional *toe* in (45) takes a PP complement and that complement is headed by *naar* ‘to’, but the latter is separated from its complement *overal* by the adjunct *op de fiets* ‘on the bike’. As a result, the adpositions form a cluster that is separated from the non-adpositional complements of the adpositions.

There are, of course, also differences between verb clusters and adposition clusters. The linear order in adposition clusters, for instance, is rigid, while the one in verb clusters is more flexible.¹³ Besides, verb clusters are much more common than adposition clusters. A count in the same sample as the one that is used in this paper yields a result of 19,074 clusters of two verbs (Augustinus 2015: 127), which is 44 times more than the total of 432 clusters of two adpositions. From a structural point of view, though, the phenomena are so much alike that it is worthwhile to explore whether a uniform account makes sense. This is the aim of the next section.

4 Analysis

There are many proposals in the literature for the treatment of verb clustering, on the one hand, and adposition stranding, on the other hand, but there are no proposals yet for the treatment of adposition clustering. The purpose of this section is to fill this gap. Starting from the observation that adposition clusters have much in common with verb clusters, we aim for an analysis that is sufficiently general to deal with both types of clusters. For this purpose we employ the framework of Head-driven Phrase Structure Grammar (HPSG), more specifically the constructional version developed in Ginzburg and Sag (2000). Since this is a monostratal framework the phenomena of stranding and clustering are not modeled in terms of movement. There is, for instance, no transformation that

¹³ A possible explanation for this difference is given in Van Eynde (2017: 90).

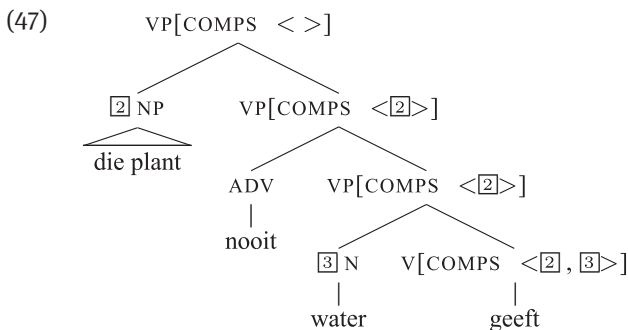
moves the complement of a verb or adposition out of its local projection into some given landing site. Instead, the non-locally realized complement is related to its head by means of the sharing of selection requirements. To show how this works, we first discuss the local realization of complements (4.1) and then the two types of non-local realization that trigger clustering, i.e. complement raising (4.2) and complement extraction (4.3). In a last step we argue why it is important to differentiate complement raising from complement extraction (4.4).

4.1 Complement realization

The subordinate clause in (46) contains a verb (*geeft* ‘gives’) which selects two complements, i.e. a direct object (*water* ‘water’) and an indirect object (*die plant* ‘that plant’).

- (46) *het schijnt dat ze die plant nooit water geeft*
 it seems that she that plant never water gives
 ‘it seems that she never waters that plant’

To model this selection HPSG employs the valence feature *COMPS*. It is assigned to all signs (words as well as phrases) and its value is a list that contains the elements which the sign selects as a complement.¹⁴ The verb *geeft* ‘gives’, for instance, has two nominals on its *COMPS* list: one for the direct object and one for the indirect object. When combined with a nominal, such as *water* ‘water’ in (46), the resulting phrase has still one nominal on its *COMPS* list. The addition of an adjunct, such as *nooit* ‘never’, does not affect the *COMPS* list, but the addition of the indirect object does, yielding a phrase with an empty *COMPS* list, as shown in (47).¹⁵



¹⁴ Technically, these elements are objects of type *synsem*, i.e. bundles of syntactic and semantic features. Not included in the *synsem* objects are phonological and pragmatic features.

¹⁵ This piecemeal addition of the complements is also proposed for German in Müller (2002).

To model this in general terms we use the implicational constraint in (48).¹⁶

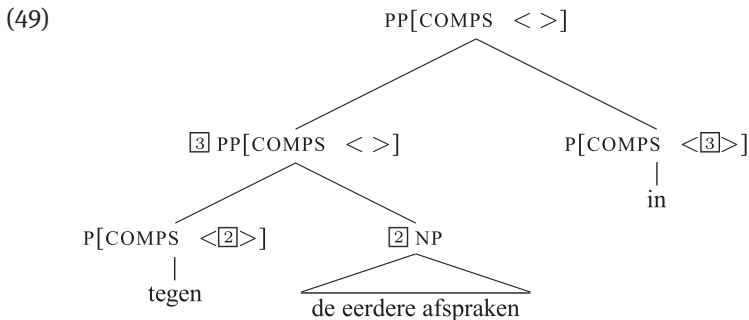
(48) **Head-Complement Phrase:**

$$\text{head-comp-phrase} \Rightarrow \left[\begin{array}{l} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} \quad \boxed{A} \\ \text{HD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} \quad \boxed{A} \oplus \langle \boxed{1} \rangle \\ \text{NONHD-DTR} \mid \text{SYNSEM} \quad \boxed{1} \text{synsem} \end{array} \right]$$

What this constraint states, is that in signs of type *head-complement-phrase*

- the syntactic and semantic properties of the non-head daughter match the requirements that the head daughter imposes ($\boxed{1}$);
- the saturated COMPS requirement is absent from the COMPS list of the mother;
- the other elements on the COMPS list, if any, are shared with the mother (\boxed{A}).

Since (48) does not impose any constraints on the part of speech of the daughters, it also subsumes phrases with a circumposition, as in (49).



The adposition *tegen* ‘against’ selects a nominal complement ($\boxed{2}$) and when that complement is added, the result is a PP with an empty COMPS list. That PP is in its turn selected by the adposition *in* ‘in’ ($\boxed{3}$) and the result is again a PP with an empty COMPS list.

A minor complication for this treatment is that complements are not always overtly realized. The direct object complement of *eaten*, for instance, is not overtly realized in *we have not eaten yet*. The canonical way to model this in HPSG is to make the NP on the COMPS list of the verb optional, as in [COMPS <(NP)>]. This intransitive use is also possible with certain adpositions. In (50), for instance, the complements of *voor* ‘for’ and *tegen* ‘against’ are not overtly realized.

¹⁶ \boxed{A} stands for any list, including the empty list. \oplus stands for the concatenation of lists.

- (50) *Wie is voor? En wie is tegen?*
 who is for? And who is against?
 ‘Who is in favor (of it)? And who is against (it)?’

Unsurprisingly, the phenomenon also affects the circumpositions, as in (51).

- (51) *als tussengerecht kwam dan de sjellik met een lepel soep* —
 P entremets came then the sjellik P a spoon soup —
overheen
 P₁.P₂
 ‘as entremets there was the sjellik with a spoonful of soup over it
 (wr-p-e-i-0000050381.p.1.s.295)

To model this we allow the NP complement of *over* to be optional. Technically, this means that its COMPS value is allowed to be the empty list, as in (52).

- (52)
-
- ```

graph TD
 A["PP[COMPS <>]"] --- B["P[COMPS <>]"]
 A --- C["P[COMPS <[]>]"]
 B --- D["over"]
 C --- E["heen"]

```

Of the 432 adposition clusters in the sample there are 12 (2.78 %) with an omitted complement (i.e. without overt complement).

A more serious complication for the treatment of complement realization is the fact that complements are not always realized in the local projection of their selector. To model this we adopt some terminology from topological field theory, as described amongst others in Engel (1970) and Höhle (1986). In that theory clauses are partitioned in two poles and three fields. The first pole, also known as the *linke Satzklammer*, is filled by the finite verb in main clauses and by the complementizer in subordinate clauses. The second pole, also known as the *rechte Satzklammer*, is filled by the verb cluster; it contains the non-finite verbs (if any); in subordinate clauses it also contains the finite verb. The space before the first pole is the *Vorfeld*, the space between the two poles is the *Mittelfeld*, and the space after the second pole is the *Nachfeld*.

Making use of these distinctions we differentiate between complement raising and complement extraction. The former concerns the realization of a complement outside of the local projection of its selector, but inside the *Mittelfeld*. The latter concerns the realization of a complement in the *Vorfeld*.

## 4.2 Complement raising

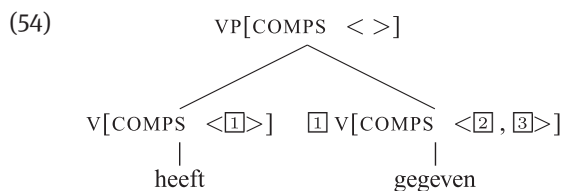
To model complement raising we take a cue from the analysis of the Dutch verb clusters in Van Eynde and Augustinus (2013).<sup>17</sup> It is briefly presented in 4.2.1 and extended to the analysis of adposition clusters in 4.2.2. Vacuous complement raising is discussed in 4.2.3. In the sample complement raising accounts for the formation of 308 of the 432 adposition clusters (71.30 %).

### 4.2.1 Complement raising out of verbal projections

If we replace the simple present *geeft* ‘gives’ in (46) with the present perfect *heeft gegeven* ‘has given’, we get a combination in which the participle is separated from its complements by the auxiliary.

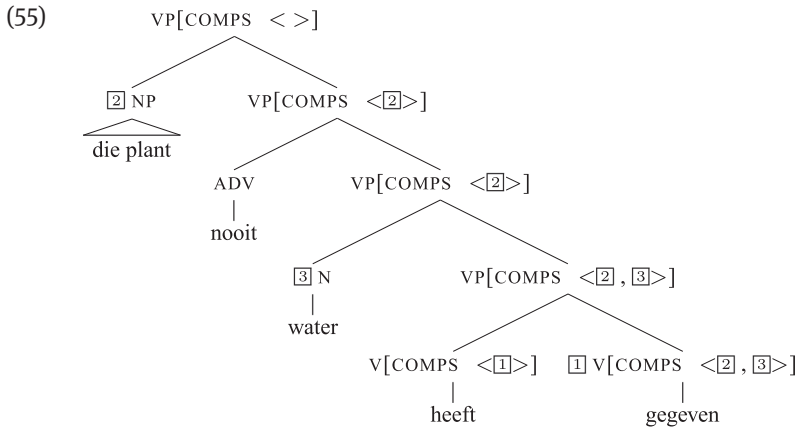
- (53) *het schijnt dat ze die plant nooit water heeft — gegeven*  
 it seems that she that plant never water has — given  
 ‘it seems that she never watered that plant’

Assuming that the auxiliary selects a participial complement, the application of the constraint in (48) yields (54).



This, however, is not what we want. Since the COMPS requirements of the participle are unsaturated, they should be propagated up the tree, till the point at which the selected nominals are added, as in (55).

<sup>17</sup> It is similar to the analysis of the German verb clusters in Hinrichs and Nakazawa (1989) and Hinrichs and Nakazawa (1994), but there are also some non-trivial differences, see Section 5.1.



To model this we add the implicational constraint in (56), quoted from Van Eynde and Augustinus (2013).

(56) **Complement Raising:**

$$\textit{headed-phrase} \Rightarrow \left[ \begin{array}{l} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} \quad \boxed{A} \oplus \boxed{B} \\ \text{HD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} \quad \boxed{B} \\ \text{NONHD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} \quad \boxed{A} \end{array} \right]$$

What (56) says, is that in signs of type *headed-phrase*, the unsaturated COMPS requirements of the non-head daughter ( $\boxed{A}$ ) are added to those which the mother inherits from the head daughter ( $\boxed{B}$ ).<sup>18</sup>

To understand the effect of this constraint, it has to be considered in tandem with the constraint on head-complement phrases in (48). While (48) has the effect of removing elements from the COMPS list, (56) has the effect of adding elements to the COMPS list. As a consequence, since a head-complement phrase is by definition a headed phrase—in HPSG parlance the former is a subtype of the latter—it is possible for a COMPS list to be extended with new members while at the same time losing other members. In fact, this is what happens in the node which immediately dominates the verb cluster in (55). Its COMPS list no longer contains the requirement for a past participle, but is extended with the nonsaturated COMPS requirements of that participle.

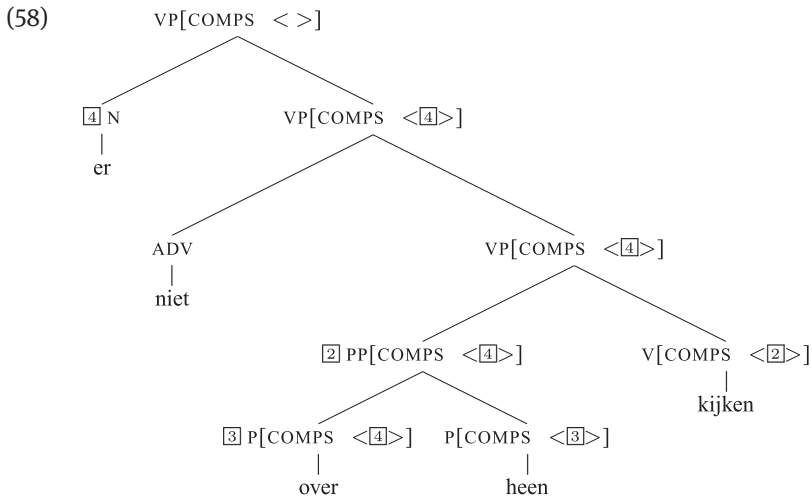
<sup>18</sup> In signs of type *non-headed phrase*, such as coordinate phrases, the COMPS list of the mother is identified with the COMPS lists of each of the conjunct daughters separately, rather than with the concatenation of those lists. In *he buys and sells cars*, for instance, the coordinate phrase *buys and sells* has the same COMPS list as its conjunct daughters *buys* and *sells*.

### 4.2.2 Complement raising out of adpositional projections

Since there are no constraints on the part of speech of the daughters in (56), the latter also models the raising out of adpositional projections, as in (57), where the R-pronoun is raised out of the PP.

(57) *ze konden er niet \_\_ over heen kijken*  
 they could it.R not \_\_ P<sub>1</sub> P<sub>2</sub> look  
 ‘they could not look over it’

Employing the constraints in (48) and (56), the analysis of (57) looks as follows:



The COMPS requirement of *kijken* (2) is immediately saturated and so is the one of *heen* (3), but the requirement of *over* for a nominal complement (4) is not. It is added to the COMPS list of the PP, inherited by the VP and propagated up the tree, just as in (55). The formation of the adposition cluster in (58) is, hence, the result of the same phenomenon as the one that triggers the formation of the verb cluster in (55).

### 4.2.3 Vacuous complement raising

An issue arises when the complement of a verbal or adpositional cluster immediately precedes its lexical selector, as in (59–60).

(59) *het schijnt dat ze toen een kaartje gestuurd heeft*  
 it seems that she then a card sent has  
 ‘it seems that she sent a card then’

(60) *de menigte die hier op af kwam*  
 the crowd that this.R P<sub>1</sub> P<sub>2</sub> came  
 ‘the crowd that was attracted by this’  
 (wr-p-e-i-0000086197p.1.s.57)

For the analysis of such combinations one can adopt a left branching structure in which the nominal complements are sisters of respectively *gestuurd* ‘sent’ and *op* ‘up’, as in (61–62).

(61) *het schijnt dat ze toen* [[een kaartje gestuurd] heeft]

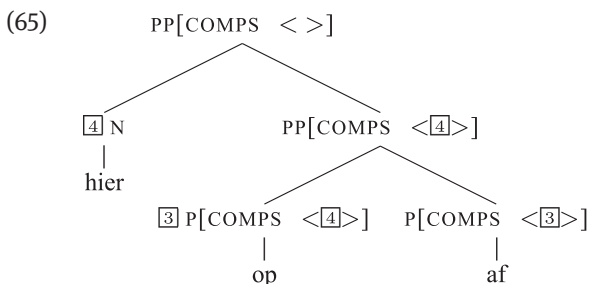
(62) *de menigte die* [[hier op] af] kwam

Alternatively, one can assign a right branching structure in which the complements are sisters of the respective clusters, as in (63–64).

(63) *het schijnt dat ze toen* [een kaartje [gestuurd heeft]]

(64) *de menigte die* [hier [op af]] kwam

The second option looks more complex, but it has the advantage of providing a more uniform account. In the case of (59), for instance, the structure in (63) is also valid if the participle follows the auxiliary, as in *een kaartje heeft gestuurd* ‘a card has sent’, while the structure in (61) is no longer applicable. By analogy, we assume that (64) is preferable to (62) for the adposition cluster. More specifically, we assume that the R-pronoun in (60) is raised out of the lower PP and realized as a complement of the cluster, as in (65).



Confirming evidence is provided by the fact the R-pronoun may scope over a conjunction of clusters, as in (66).

- (66) *of je [daar [onder door of over heen]] loopt*  
 whether you that.R P<sub>1</sub> P<sub>2</sub> or P<sub>1</sub> P<sub>2</sub> run  
 ‘whether you run under or over that’

In principle, it would be possible to treat the R-pronoun as raised out of the higher PP as well and, hence, as a dependent of the verbal projection. This, however, is neither necessary nor desirable, since the combination of the R-pronoun and the adposition cluster is treated as a constituent in the case of extraction, as illustrated in (67).

- (67) *[daar tussendoor] zongen Joan la Barbara en S. Narucki Cage’s*  
 that.R P<sub>1</sub>.P<sub>2</sub> sang Joan la Barbara and S. Narucki Cage’s  
*liederen*  
 songs  
 ‘In between Joan la Barbara and Susan Narucki sang Cage’s songs’  
 (wr-p-p-h-0000000103.p.6.s.4)

It is also treated as a constituent in the absolute *met*-construction in (68).

- (68) *De zuilen bestaan uit beton met [daar om heen] een*  
 the columns consist of concrete with that.R P<sub>1</sub> P<sub>2</sub> a  
*laag brons*  
 layer bronze  
 ‘The columns consist of concrete with a layer of bronze around that’  
 (wr-p-e-i-0000050211.p.1.s.106)

Of the 308 instances of complement raising in the sample, 156 concern vacuous raising.

### 4.3 Complement extraction

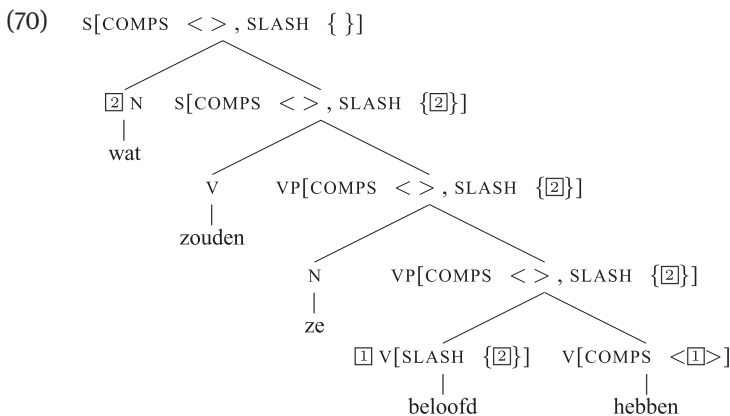
To model complement extraction we adopt the treatment of unbounded dependencies in Ginzburg and Sag (2000: ch. 5). We first discuss its role in the formation of clusters, both verbal and adpositional ones (5.3.1), and then treat the phenomenon of topic drop (5.3.2). Of the 432 adposition clusters in the sample 112 (25.92 %) result from complement extraction.

### 4.3.1 Complement extraction out of verbal and adpositional projections

An example of complement extraction is given in (69).

- (69) *wat zouden ze — beloofd hebben?*  
 what would they — promised have?  
 ‘what would they have promised?’

The direct object of *beloofd* ‘promised’ is not realized in its local projection. Instead, it appears in the *Vorfeld*, preceding the first pole, which is filled by the finite auxiliary *zouden* ‘would’. To relate the extracted complement to its selector Ginzburg and Sag (2000) does not rely on the sharing of COMPS values. Instead, the COMPS requirement that is not locally saturated is subtracted from the COMPS list of the selector, stored in another feature, called SLASH, and propagated up the tree, as in (70).<sup>19</sup>



The subtraction is modeled in terms of the Argument Realization Principle (Ginzburg and Sag 2000: 171).<sup>20</sup>

<sup>19</sup> Empirical evidence for treating extraction differently from raising will be given in section 4.4.

<sup>20</sup> The constraint will be slightly modified in section 4.4.2.  $\ominus$  stands for list subtraction. Besides SUBJ and COMPS, Ginzburg and Sag (2000) uses a third valence feature, called SPR, for the selection of specifiers. Nouns, for instance, are claimed to select a determiner as their specifier. We do not use this feature, since we adopt the functor treatment of determiners, as developed in Van Eynde (2006). In that treatment, it is the specifier which selects its head sister, rather than the head which selects its specifier.



(71) **Argument Realization Principle:** (preliminary version)

$$\text{word} \Rightarrow \left[ \begin{array}{l} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \left[ \begin{array}{l} \text{SUBJ} \quad \boxed{A} \\ \text{COMPS} \quad \boxed{B} \ominus \text{list}(\text{gap-synsem}) \end{array} \right] \\ \text{ARG-ST} \quad \boxed{A} \oplus \boxed{B} \end{array} \right]$$

The subtracted COMPS requirement is an object of type *gap-synsem* and such objects have the defining property that their LOCAL value, which contains their syntactic and semantic properties, is added to their SLASH set (Ginzburg and Sag 2000: 170).

(72) **Gaps:**

$$\text{gap-synsem} \Rightarrow \left[ \begin{array}{l} \text{LOC} \quad \boxed{1} \\ \text{SLASH} \quad \{\boxed{1}\} \end{array} \right]$$

In other words, if the selector requires, say, an accusative NP complement, then this requirement is stored in the SLASH value. To model its propagation we assume that the SLASH set of the mother equals the union of the SLASH sets of the daughters. This stops when the slashed clause (*zouden ze beloofd hebben*) is combined with the extracted complement (*wat*). This is modeled by the constraint on phrases of type head-filler in (73), adapted from Ginzburg and Sag (2000: 174).<sup>21</sup>

(73) **Head-Filler Phrase:**

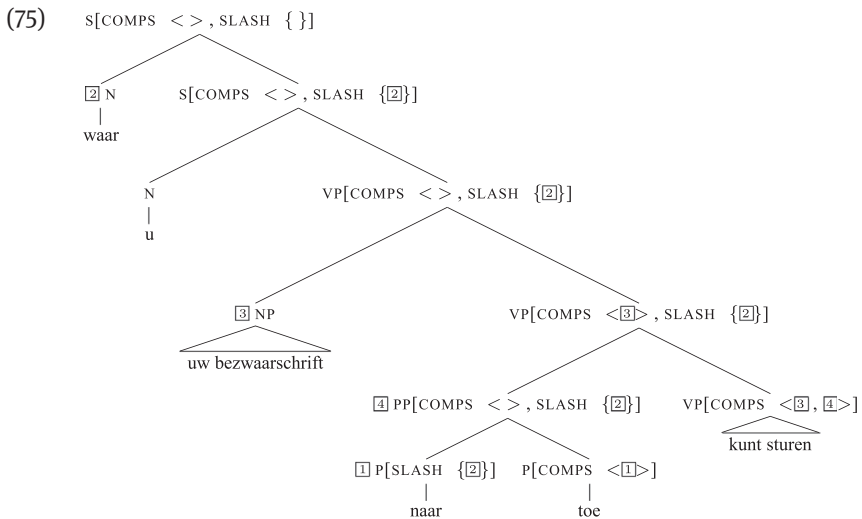
$$\text{head-filler-phrase} \Rightarrow \left[ \begin{array}{l} \text{SYNSEM} \mid \text{SLASH} \quad \boxed{\Sigma} \\ \text{HD-DTR} \left[ \begin{array}{l} \text{SYNSEM} \left[ \begin{array}{l} \text{LOC} \mid \text{CAT} \mid \text{HEAD} \quad \textit{verb} \\ \text{SLASH} \quad \{\boxed{1}\} \uplus \boxed{\Sigma} \end{array} \right] \end{array} \right] \\ \text{NONHD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \quad \boxed{1} \textit{local} \end{array} \right]$$

In signs of type *head-filler-phrase*, the head daughter is required to be a verbal projection whose SLASH set contains an element of type *local* that matches the LOC value of the non-head daughter. This checks whether the requirements which the selector of the extracted element imposes are met. In (70), for instance, the extracted element must be an accusative NP.

Extraction out of adpositional phrases can be modeled in the same way, as illustrated for the circumpositional PP in (74–75).

<sup>21</sup>  $\uplus$  stands for disjoint set union.

- (74) *waar u uw bezwaarschrift \_\_ naar toe kunt sturen*  
 what.R you your complaint \_\_ P<sub>1</sub> P<sub>2</sub> can send  
 ‘where you can send your complaint to’  
 (wr-p-e-i-0000049645.p.1.s.156)



The adposition *naar* saturates the COMPS requirement of *toe* (1), but its own requirement for a nominal is subtracted from its COMPS list, added to its SLASH set, and propagated up the tree till the point where the R-pronoun is added.

### 4.3.2 Topic drop

Extracted complements are not always overtly realized. The phenomenon is known as topic drop and is exemplified by such elliptical clauses as *doen we* ‘do we’ and *ken ik niet* ‘know I not’, where the omitted topic is a pronoun like *dat* ‘that’. It also affects the extracted complements of circumpositions, as in (76).

- (76) *ja goed. maar \_\_ kom ik ook wel \_\_ overheen denk je*  
 yes ok. but \_\_ come I also \_\_ P<sub>1</sub>.P<sub>2</sub> think you  
*niet?*  
 not?  
 ‘Yes ok, but I’ll get over it, don’t you think so?’  
 (fnc008021\_174)

The omitted topic is understood to be *daar* ‘that.R’ and the omission is optional: Adding it in the position of the first dash yields a well-formed combination.

It is worth stressing that topic drop is distinct from complement omission. There is, first of all, a clear stylistic difference. While complement omission occurs both in spoken and in written Dutch, including formal registers, topic drop is typical of informal spoken Dutch: All instances in the sample are from the spoken Dutch treebank (15 from the Netherlands and 1 from Belgium). Another difference concerns the omitted material: While the omitted complement of intransitively used adpositions is *er* ‘it.R’, the omitted element in the case of topic drop is *daar* ‘that.R’.

To model topic drop, we treat it as a filler-gap mismatch: The requirement of the verb or adposition for a nominal complement is subtracted from the COMPS list, added to the SLASH set and propagated up the tree in the usual fashion. The only unusual bit is that the filler is not overtly realized. Another instance of a filler-gap mismatch in the sample is (77).

- (77) *naar Euro Disney zou 'k ook nog wel 'ns een keertje* \_\_\_  
 to Euro Disney would I also still once a time \_\_\_  
*naartoe willen*  
 P<sub>1</sub>.P<sub>2</sub> want  
 ‘to Euro Disney I’d like to go to some time as well’  
 (fna000784\_252)

While the extracted complement is an NP, the filler is a PP. Just like topic drop, this mismatch is stylistically marked. In fact, most speakers find it ill-formed.

#### 4.4 Constraints on complement raising and complement extraction

There is a lot that complement raising and complement extraction have in common. In both cases there is a complement that is not realized in the local projection of its selector and in both cases there is a device of sharing selection requirements in order to relate the non-locally realized complement to its selector. The only difference, it seems, is the position of the complement: While raised complements appear in the *Mittelfeld*, extracted complements appear in the *Vorfeld*. The purpose of the section now is to show that this is not the only difference. More specifically, we will show that there are complements which can be raised but not extracted (4.4.1) and we will argue that raising and extraction are subject to different constraints (4.4.2).

#### 4.4.1 Major vs. minor

Most of the complements that can be raised can also be extracted, and vice versa, but there is an important exception in the case of the pronouns. For many of its pronouns Dutch has both a full form and a reduced form. The former have a clear vowel, as in *jou* ‘you’, and *mij* ‘me’, while the latter usually have a mute vowel, as in *je* ‘you’ and *me* ‘me’. Both the full and the reduced forms can be raised out of the verb cluster, as in (78), but as the contrast in (79) shows, only the full forms can be extracted.

(78) *ze hadden jou/je meteen moeten* — *ontslaan*  
 they had you immediately must — fire  
 ‘they should have fired you immediately’

(79) *jou/\*je hadden ze meteen moeten* — *ontslaan*  
 you had they immediately must — fire  
 ‘they should have fired you immediately’

This correlates with some other differences: While the full forms can take dependents, such as *alleen* ‘alone’, and can be conjoined, the reduced forms cannot.

(80) *we hebben alleen jou/\*je gezien*  
 we have only you seen  
 ‘we saw only you’

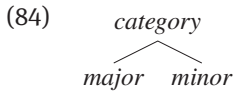
(81) *Peter stond tussen jou/\*je en mij/\*me*  
 Peter stood between you and me  
 ‘Peter stood between you and me’

The distinction is also relevant for the R-pronouns, where the full form *daar* contrast with the reduced forms *d’r* and *er*. As expected, both can be raised, as in (82), but only the full form can be extracted, as shown in (83).

(82) *u kunt daar/er uw bezwaarschrift* — *naar toe sturen*  
 you can that.R/it.R your complaint — P<sub>1</sub> P<sub>2</sub> send  
 ‘you can send your complaint there.’

(83) *daar/\*er kunt u uw bezwaarschrift* — *naar toe sturen*  
 that.R/\*it.R can you your complaint — P<sub>1</sub> P<sub>2</sub> send  
 ‘that you can send your complaint to’

It may be worth stressing that the distinction concerns a syntactic property, rather than a phonological one, for while it is true that most of the reduced forms have a mute vowel as their nucleus, this is not always the case. The reflexive pronoun *zich*, for instance, has a clear vowel, but shows the defining characteristics of reduced forms: When topicalized, modified or conjoined, it must be replaced by the full form *zichzelf*, see Van Eynde (1999: 146). For this reason we make the distinction in the CATEGORY value of the signs, rather than in the PHONOLOGY value. More specifically, we assume that the values of the CATEGORY feature come in two types, which we will call major and minor, after Van Eynde (1999: 141).



The distinction is orthogonal to the part-of-speech distinction. Technically, this follows from the fact that part-of-speech labels are values of the HEAD feature, rather than of the CATEGORY feature. Empirically, it is motivated by the fact that the distinction between major and minor is not only relevant for various types of pronouns (personal, reflexive, demonstrative, ...), but also for adpositions, as shown in Van Eynde (2004), and for determiners, as shown in Van Eynde (2006). Employing this distinction, we can formulate the relevant constraint as a restriction on the arguments which are subtracted from the COMPS list.

(85) **Argument Realization Principle:** (final version)

$$\textit{word} \Rightarrow \left[ \begin{array}{l} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \\ \text{ARG-ST} \quad \boxed{A} \oplus \boxed{B} \end{array} \left[ \begin{array}{l} \text{SUBJ} \quad \boxed{A} \\ \text{COMPS} \quad \boxed{B} \ominus \textit{list} \left( \left[ \begin{array}{l} \textit{gap-synsem} \\ \text{LOC} \mid \text{CAT} \quad \textit{major} \end{array} \right] \right) \end{array} \right] \right]$$

(85) requires the arguments which are subtracted from the COMPS list to be major. Minor signs can, hence, not be extracted. Since this constraint does not hold for complement raising it provides empirical evidence for the differentiation between complement raising and complement extraction.

When assessing the relevance of this constraint for other languages, it should be borne in mind that the major/minor distinction is subject to cross-linguistic variation. The fact that a word is minor in one language does not necessarily imply that its translational equivalents are minor as well. The Dutch pronoun *het* 'it', for instance, is minor, but its English equivalent is not, as demonstrated by the fact that it can be conjoined and modified, as in (86).

- (86) you might be tempted to read it and it alone, fanatically, ...  
 (quote on the cover of Don DeLillo's *Underworld*)

It should also be borne in mind that the distinction does not coincide with the one between tonic and clitic pronouns: While minor signs are independent lexical units, clitics are often incorporated in another element, usually a verb.

#### 4.4.2 Initial vs. final

Another reason for differentiating between complement raising and complement extraction is that there are languages which allow complement extraction but not complement raising. One of them is English. That it allows complement extraction is well known. (87) provides examples of extraction out of verbal and adpositional projections.

- (87) that topic I never want to discuss \_\_ again

- (88) what did you say she was speaking about \_\_ ?

Complement raising, however, does not seem to exist in English. In fact, it is explicitly ruled out by the Empty Comps Constraint, as spelled out in Ginzburg and Sag (2000: 33).<sup>22</sup>

- (89) **Empty Comps Constraint:**

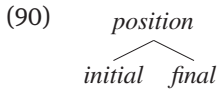
$$phrase \Rightarrow [SYNSEM | LOC | CAT | COMPS \langle \rangle ]$$

This constraint makes it impossible for a phrase to inherit unsaturated COMPS requirements of its daughters.

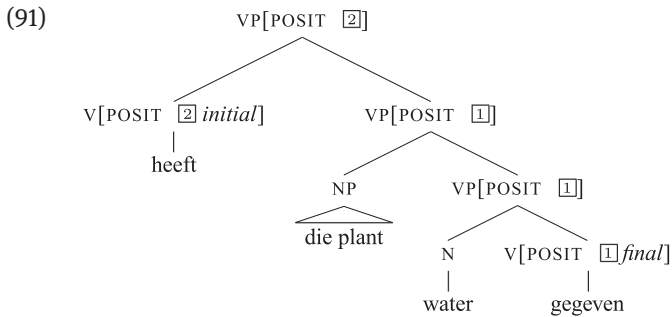
For Dutch, which does allow complement raising, the constraint in (89) is obviously too strong, but that does not mean that its complements are allowed to be raised anywhere. For a start, they must be realized in the *Mittelfeld*. The first pole is, hence, a barrier for complement raising. To model this we need a way to define the first pole. Given that it contains either a complementizer or a verb, we

<sup>22</sup> It might make sense to restrict this to *headed* phrases, since non-headed phrases may be unsaturated for COMPS. The bracketed coordinate phrase in *he [buys and sells] cars*, for instance, has a non-empty COMPS list.

introduce a feature that is assigned to both. We call it *POSITION* and declare its values to be *initial* and *final*.<sup>23</sup>



The complementizers are invariably initial, but the verbs show more variation: The non-finite forms are final, the imperative forms are initial and the other finite forms are underspecified. The latter's *POSITION* value is resolved to *initial* if they occur in the first pole and to *final* if they occur in the second pole. Assuming that *POSITION* is a *HEAD* feature, its value is propagated throughout the local projection, as in (91).



The non-finite *gegeven* is the head of the VPs *water gegeven* and *die plant water gegeven*, and hence shares its *POSITION* value (1) with them. The finite auxiliary, by contrast, is the head of the VP *heeft die plant water gegeven*, and shares its *POSITION* value (2) with the top node.

Employing the *POSITION* feature, the barrier status of the first pole for complement raising can be formulated as follows:

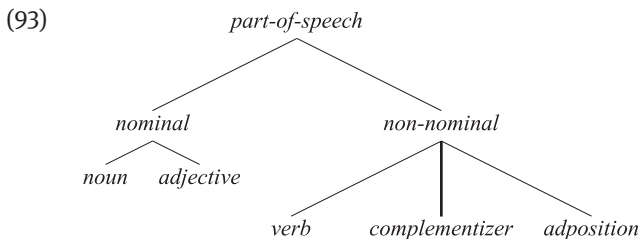
(92) **Complement Raising Barrier:**

$$\left[ \begin{array}{l} \textit{headed-phrase} \\ \text{SYNSEM | LOC | CAT | HEAD | POSIT } \textit{initial} \end{array} \right] \Rightarrow \left[ \begin{array}{l} \text{SYNSEM | LOC | CAT | COMPS } \langle \ \rangle \end{array} \right]$$

<sup>23</sup> The assumption that complementizers share properties with verbs is also made for other languages. Höhle (1997) provides evidence for German and Ginzburg and Sag (2000: 23) for English: “the part of speech types (the values of the feature *HEAD*) associated with verbs and complementizers are subtypes of a common supertype” (p. 23).

What (92) says is that a headed phrase with the POSITION value *initial* must have an empty COMPS list. An example is the top node in (91). The requirement that such phrases have an empty COMPS list implies that complements cannot be raised out of a v-initial VP, nor out of a CP. Raising out of a v-final VP, however, is allowed. Formally, (92) is similar to the Empty Comps Constraint in (89). The right hand side of the implication is, in fact, identical. The difference concerns the left hand side.

Intriguingly, the contrast between v-initial and v-final VPs is also relevant for PPs. Notice, for a start, that there are adpositions which invariably precede their complement, such as *met* ‘with’ and *tot* ‘till’,<sup>24</sup> as well as adpositions which invariably follow their complement, such as *af*, *mee* and *toe*. Besides, there are adpositions which are underspecified in this respect, such as *over* ‘about’ and *aan* ‘on’. This strongly suggests that the distinction is relevant for adpositions too. Acting on that hint, let us assume that the POSITION value is also assigned to adpositions and their projections. To model this we make a distinction between nominal and non-nominal parts of speech, where the latter comprise the verbs, the complementizers and the adpositions, and declare the POSITION feature for the non-nominal parts of speech.<sup>25</sup>



(94) *non-nominal* : [POSITION *position*]

Given the constraint in (92) it follows that not only CPs and v-initial VPs are barriers for complement raising, but also P-initial PPs. This implication turns out to be correct. To show this let us compare (95–96) with (97–98).

(95) *ik heb toen [een boek daar over] gelezen*  
 I have then a book that.R about read  
 ‘I read a book about that’

<sup>24</sup> Zwarts (1997: 1094–1095) lists no less than 57 of these adpositions.

<sup>25</sup> The notion of *non-nominal part of speech* is inspired by x-bar theory, in which the lexical categories are analyzed in terms of the boolean features N and V. In this analysis nouns are [+N, -V], adjectives are [+N, +V], verbs are [-N, +V], and adpositions are [-N, -V], see Chomsky (1970) and Jackendoff (1977).



(96) *ik heb daar toen [een boek \_\_ over] gelezen*

I have that.R then a book \_\_ about read

‘I read a book about that’

(97) *ik heb toen [aan een boek daar over] meegewerkt*

I have then to a book that.R about contributed

‘I contributed to a book about that.’

(98) \* *ik heb daar toen [aan een boek \_\_ over] meegewerkt*

\* I have that.R then to a book \_\_ about contributed

Both in (95) and (97), the POSITION value of *over* is resolved to *final*, since it is preceded by its complement *daar* ‘that.R’. By contrast, the POSITION value of *aan* in (97) is resolved to *initial*, since it precedes its NP complement. Given that P-initial PPs are a barrier for complement raising, this accounts for the fact that *daar* can be raised out of the PP in (95), but not out of the PP in (97).

Turning now to complement extraction, it is clear that it is not constrained by (92). Neither CPS nor V-initial VPS are barriers for extraction, as illustrated in (99), where the extracted complement is separated from its selector by two complementizers and the V-initial verb *denk* ‘think’.

(99) *wat denk je dat ze zei dat ze wil \_\_ kopen?*

what think you that she said that she wants \_\_ buy?

‘what do you think she said she wants to buy?’

P-initial PPs, however, do block extraction. This holds both for PPs which are headed by an inherently initial adposition, such as *met* ‘with’, and for PPs which are headed by an adposition whose POSITION value is resolved to *initial*, such as *aan* in (101).

(100) \* *daar heb ik toen een deur met \_\_ geverfd*

\* that.R have I then a door with \_\_ painted

(101) \* *waar heb jij toen [aan een boek \_\_ over] meegewerkt*

\* what.R have you then to a book \_\_ about contributed

To model this we assume that adpositions whose POSITION value is of type *initial* must have an empty SLASH set.<sup>26</sup>

<sup>26</sup> This constraint does not hold for English, as is clear from the wellformedness of (88).

(102) **Complement Extraction Barrier:**

$$\left[ \begin{array}{l} \textit{sign} \\ \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} \left[ \begin{array}{l} \textit{adposition} \\ \text{POSITION} \textit{ initial} \end{array} \right] \end{array} \right] \Rightarrow \left[ \text{SYNSEM} \mid \text{SLASH} \left\{ \right\} \right]$$

This constraint applies both to adpositions and their projections. It is similar to a constraint that is proposed for the Dutch adpositions in Tseng (2005): “we can say that only [+post] Ps can put their complement in SLASH, or equivalently, that all the SLASH set [sic] of all PREpositions must be empty.”

Returning to the main topic of this section, the fact that the constraint on complement raising in (92) differs from the constraint on complement extraction in (102) provides further evidence for differentiating the two phenomena. Moreover, zooming out and taking a broader perspective, it has long been recognized in generative grammar at large that scrambling and extraction are sufficiently different to require different treatments.

#### 4.5 Summing up

This section has shown that the factors which trigger the formation of adposition clusters are the same as those which trigger the formation of verb clusters, i.e. complement raising and complement extraction. Both phenomena have been given a fully explicit analysis which is cast in the notation of HPSG. Besides, the differentiation between complement raising and complement extraction has been backed up with empirical evidence.

## 5 A comparison with other monostratal treatments

The previous section has provided an analysis of the phenomena of clustering and stranding which is applicable to both verbs and adpositions. In this respect, it departs from the common practice in HPSG to treat clustering and stranding as unrelated, with the former being confined to verbs and the latter to adpositions. Possibly as a consequence of this assumption of unrelatedness, the treatments are very different. To illustrate this we start from the scheme in Table 4. While HPSG treatments of verb clustering commonly ignore the distinction between complement raising and subject raising (Section 5.1), HPSG treatments of adposition stranding commonly ignore the distinction between complement raising and complement extraction (Section 5.2). We will briefly present these proposals and

**Table 4:** Raising vs. extraction.

|            | Subject            | Complement            |
|------------|--------------------|-----------------------|
| Raising    | Subject raising    | Complement raising    |
| Extraction | Subject extraction | Complement extraction |

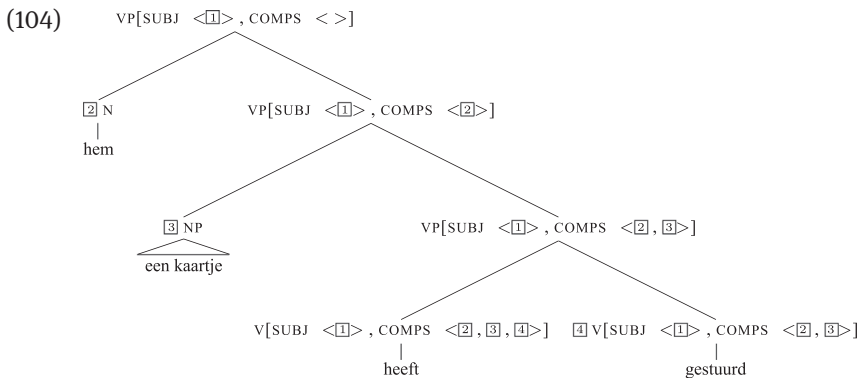
show why it is preferable to treat complement raising as distinct from both subject raising and complement extraction.

## 5.1 Verb clustering and generalized raising

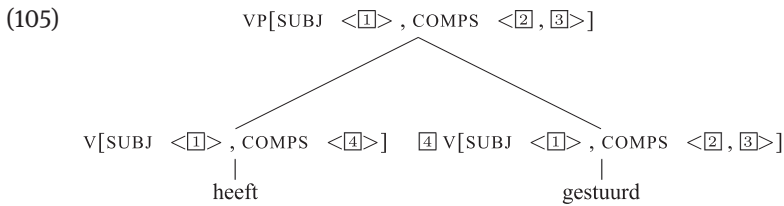
The canonical way to model verb clusters in HPSG is based on generalized raising. It was originally proposed for a treatment of German verb clusters in Hinrichs and Nakazawa (1989) and Hinrichs and Nakazawa (1994), and got adopted and adapted by various authors to deal with similar phenomena in other languages, such as Dutch verb clusters (Bouma and van Noord 1998) and clitic climbing in French (Abeillé et al. 1998) and Italian (Monachesi 1998). As an illustration of how it works, let us take (103).

- (103) *het schijnt dat ze hem een kaartje heeft gestuurd*  
 it seems that she him a card has sent  
 ‘It seems that she sent him a card’

To model the dependencies in this clause the auxiliary *heeft* ‘has’ is claimed to inherit the argument requirements of its participial complement. These include both the subject and the complement requirements. The auxiliary in (103), hence, selects four arguments: its own participial complement, augmented with the three arguments of that participle.



In this analysis subject raising and complement raising are treated in the same lexicalist head-driven way. In our treatment, by contrast, the subject requirement (1) of the participle is inherited by the perfect auxiliary, but its complement requirements (2 and 3) are propagated directly to the cluster and beyond. Technically, the difference only shows in the COMPS value of the auxiliary. The rest of the structure is identical to (104).



Conceptually, the difference is more important. What it intends to capture is the assumption that subject raising is indeed amenable to the lexicalist head-driven analysis that is familiar from the treatment of subject raising in English, but that complement raising requires another approach.

There are at least four arguments in favor of this differentiation. First, subject raising is a phenomenon that affects a limited number of lexical signs, more specifically the subject-to-subject raisers (modals, temporal and aspectual auxiliaries, ...) and the subject-to-object raisers (perception verbs, causatives, ...). These include mostly verbs and—in some languages—predicative adjectives, such as *likely* and *bound* Ginzburg and Sag (2000: 21). Complement raising, by contrast, is a much more general phenomenon. In the languages which allow it, such as Dutch and German, it occurs in all sorts of verbal and adpositional constructions, as illustrated in Section 4.2.

Second, subject raising and complement raising are mutually independent, in the sense that there are instances of complement raising which do not involve subject raising, and that there are instances of subject raising which do not involve complement raising. Starting with the former, subject control verbs, such as the Dutch *willen* ‘want’ and *proberen* ‘try’, obviously do not belong to the subject raising verbs, but they do allow complement raising, as shown in (106–107), where the bracketed complements of *stoppen* ‘stop’ and *bellen* ‘call’ are raised out of the infinitival VP complements of *willen* ‘want’ and *geprobeerd* ‘tried’.<sup>27</sup>

<sup>27</sup> The distinction between subject control verbs and subject raising verbs is discussed amongst others in Pollard and Sag (1994: 132–145) and Sag et al. (2003: 364–376).

- (106) *dat hij [het bloedvergieten] niet had willen \_\_ stoppen*  
 that he the bloodshed not had want \_\_ stop  
 ‘that he had not wanted to stop the bloodshed’  
 (dpc-ind-001648-nl-sen.p.19.s.6)

- (107) *ik heb [’r] geprobeerd \_\_ te bellen maar ...*  
 I have her tried \_\_ to call but ...  
 ‘I tried to call her but ...’  
 (fna000583\_351)

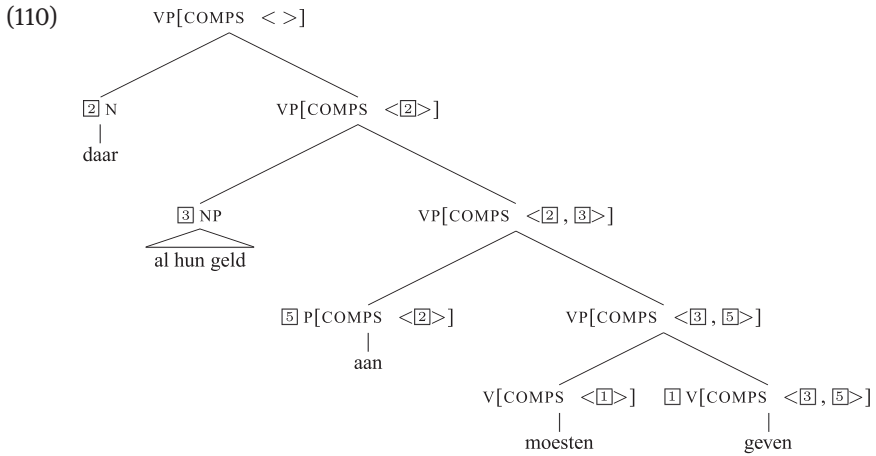
Similarly, adpositions which introduce a PP-complement are assumed to lack a SUBJ requirement in HPSG and, hence, do not qualify as subject raisers, but they do allow complement raising, as illustrated in Section 4.2. Conversely, there are instances of subject raising which do not involve complement raising. English, for instance, has subject raising verbs and adjectives, but does not allow complement raising, and in languages which allow both, such as Dutch, one can have subject raising without complement raising, as in (108), where the nonreferential subject of the infinitival complement of *lijkt* ‘seems’ is raised, but where the complement of the infinitive is realized in situ.

- (108) *er lijkt een probleem te zijn met de watervoorziening*  
 there seems a problem to be with the water.supply  
 ‘there seems to be a problem with the water supply’

Third, the non-lexicalist treatment of complement raising is also appropriate for the analysis of clauses which contain both a stranded adposition and a verb cluster, as in (109).

- (109) *dat ze daar al hun geld \_\_ aan moesten \_\_ geven*  
 that they that.R all their money \_\_ to must \_\_ give  
 ‘that they had to give all their money to that’

The unsaturated COMPS requirement of *aan* ‘to’ is added to the COMPS value of the VP, in conformity with the constraint on complement raising in (56). No extra stipulations are needed to model this.

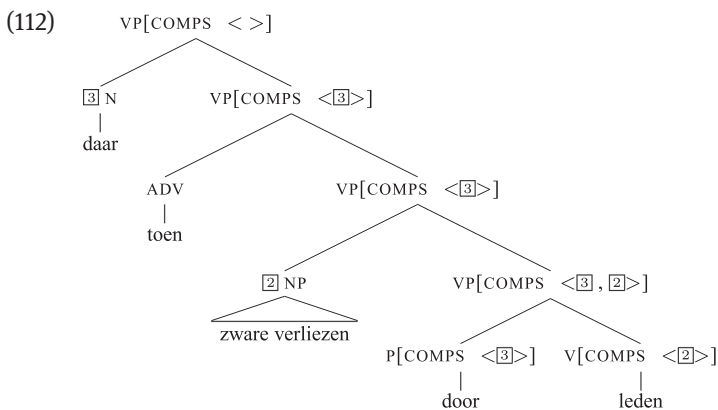


In the lexicalist treatment of generalized raising, though, there is a problem, since the unsaturated COMPS requirement of *aan* has to be included in the COMPS list of the modal, even though the modal is not a sister of the adposition.

Fourth, the non-lexicalist treatment of complement raising also deals with raising out of PP adjuncts, as in (111).

- (111) ...*dat we daar toen zware verliezen* \_\_ *door leden*  
 ...that we that.R then heavy losses \_\_ by suffered  
 ‘... that we suffered heavy losses because of that’

The unsaturated COMPS requirement of the adposition (3) is added to that of the verb (2), yielding a VP that selects two nominal complements, as in (112).



This is possible since the constraint on complement raising applies to all headed phrases. In the generalized raising analysis, by contrast, there is a problem, since adjuncts are not lexically selected by their head sister.

Summing up, there are at least four arguments for differentiating subject raising from complement raising. In fact, this list could be extended. Van Eynde and Augustinus (2013), for instance, provides two further arguments, relating to the interaction with binding and passivization, but we leave these aside, since the presentation would lead us too far astray from the central topic of this paper.

## 5.2 Adposition stranding and extraction

Treatments of adposition stranding in HPSG tend to ignore the distinction between complement raising and complement extraction. Rentier (1993) and Müller (1995), for instance, take it for granted that adposition stranding invariably results from extraction and focus on how it can be implemented. Bouma (2000) does consider the option of differentiating between complement extraction and argument inheritance (which includes complement raising), but then dismisses it and argues for a uniform extraction analysis: “Although argument inheritance plays an important role in the syntax of Dutch verb clusters an approach based on argument inheritance seems highly unlikely for R-pronouns” (Bouma 2000: 69). The dismissal is based on four arguments which we will present and discuss one by one. This discussion is an update of Van Eynde and Augustinus (2014).

First of all, prepositions which do not allow extraction (such as *met*) cannot be associated with an R-pronoun in the *Mittelfeld* either. If two different mechanisms are used to account for these two phenomena, such generalizations are easily lost. (Bouma 2000: 69)

Our treatment indeed employs different devices for complement raising and complement extraction, but this does not cause a loss of generalization, since the constraints on these devices have a different range of application: While the constraint on complement raising in (92) extends to head-initial VPs and CPs, the constraint on complement extraction in (102) does not.

Second, as argument inheritance normally involves the composition of two COMPS lists, R-pronouns would have to be allowed on COMPS, even though they can, apart from a few exceptional cases, never appear in a position following the preposition. (Bouma 2000: 69)

Our treatment indeed allows R-pronouns on COMPS lists. That they never appear in a position following the adposition does not by itself imply that this is an illicit assumption, since many of the Dutch adpositions allow their complement to precede them, as shown in (113).

- (113) *dat hij achteruit [de garage in] reed*  
 that he backward the garage in drove  
 ‘that he drove into the garage backwards’

In fact, the inherently final adpositions, such as *af*, *mee* and *toe*, even require their complement to precede them. Yet, this is not by itself a valid argument for not allowing the NP or PP which they select in their COMPS list. Indeed, given that Dutch has V-final VPs and A-final APs, as shown in (114) and (115), it is only natural to assume that it also has P-final PPs.

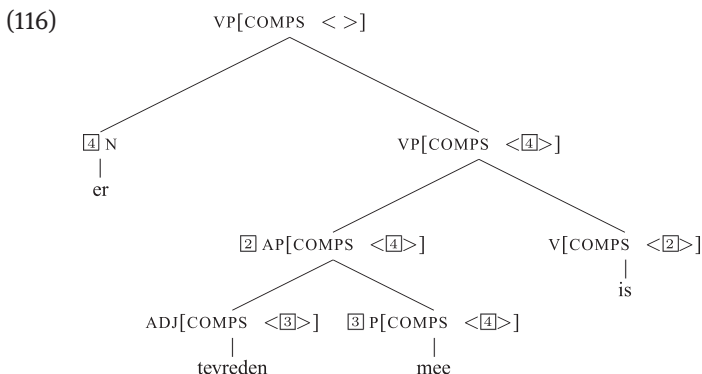
- (114) *dat hij [een grote mond heeft]*  
 that he a big mouth has  
 ‘that he has a big mouth’

- (115) *dat hij [haar fratsen beu] is*  
 that he her antics fed-up is  
 ‘that he is fed up with her antics’

There is, hence, nothing anomalous about the presence of R-pronouns on COMPS lists of adpositions.

Third, the set of argument inheritance verbs must now not only contain auxiliaries and modals, but all verbs which select a (prepositional) complement. Examples such as *Kim is er tevreden mee* ‘Kim is happy about it’ introduce further complications for an argument inheritance approach, as it suggests that predicative adjectives and nouns must be argument inheritors as well. (Bouma 2000: 69)

This objection is indeed valid for the generalized raising treatment, but not for our treatment, since we do not add the unsaturated COMPS requirement of the adposition to those of its head sister. Instead, it is propagated directly to the AP and the VP, as illustrated in (116).





The unsaturated COMPS requirement of *mee* is shared with the AP *tevreden mee* and the VP *tevreden mee is*, but it is not shared with the adjective nor with the copula.

Finally, ... in an argument inheritance approach, the relationship between valence and syntactically realized arguments has to be one-on-one, and thus there is no room for amalgamation of syntactic functions. (Bouma 2000: 69)

An example of amalgamation is (117).

- (117) *Er komt een prins \_\_\_ in voor*  
 there comes a prince \_\_\_ in SVP  
 ‘There is a prince in it.’

In Gosse Bouma’s analysis *er* simultaneously fulfills two syntactic functions in (117). It is both the expletive subject of the clause and the complement of the stranded adposition. This amalgamation, he claims, is impossible to model in terms of argument inheritance, since that device does not allow for a one-to-many relation between syntactically realized arguments and valence. While this may be a valid criticism for the argument inheritance approach, it is not necessarily a problem for our treatment.

For a start, notice that in our treatment *er* has only one syntactic function. It is the expletive subject of the clause, but it is not the complement of the stranded adposition. In fact, it cannot be the complement of the adposition, since the constraints on raising and extraction do not allow this. More specifically, the complement raising barrier in (92) blocks the raising out of a v-initial VP, and the constraint on extraction in (85) blocks the extraction of minor complements. (117) is, hence, not an instance of amalgamation. Instead, we assume that it is an instance of complement omission.<sup>28</sup> Evidence for this assumption is provided by the fact that the subject *er* is expletive, while the non-realized complement of the adposition is not. This is a problem for the amalgamation analysis since it has the undesirable consequence that the same word is required to be expletive and referential. The complement omission analysis does not have this problem, since *er* has only one function. Another piece of evidence is provided by infinitival clauses that are introduced by the complementizer *om* ‘for’, as in (118).

- (118) *dat vind ik nog wel leuk om \_\_\_ naar toe te gaan*  
 that find I still nice \_\_\_ P<sub>1</sub> P<sub>2</sub> to go  
 ‘that I would like to go to’  
 (fna000714\_240)

<sup>28</sup> A similar assumption is made in the transformational treatment of Bennis (1986).

The unrealized complement of the stranded circumposition is understood to be *er* ‘it.R’, but there is no overt *er* elsewhere in the sentence with which it could be claimed to be amalgamated. The omission treatment does not have this problem.

Taking stock, Bouma (2000) lists four arguments against the differentiation between argument inheritance and complement extraction, but closer scrutiny reveals that none of them sticks. Conversely, the discussion of the constraints on complement raising and complement extraction in Section 4.4 has shown that these constraints are sufficiently different to motivate a treatment which differentiates complement raising from complement extraction.

## 6 Conclusion

This paper has accomplished three things. First, it has demonstrated that Dutch not only has verb clusters, but also adposition clusters. These are stranded circumpositions which are headed by a postposition and which have two or three members in a rigid linear order. That part is based on Van Eynde (2017). Second, it has shown that stranding and clustering, both of verbs and adpositions, are the result of the same syntactic processes, i.e. complement raising and complement extraction. The analysis of these phenomena is cast in the notation of Head-driven Phrase Structure Grammar, more specifically in the constructional version of Ginzburg and Sag (2000). Third, it has argued that the resulting treatment is more adequate than the existing treatments of verb clustering and adposition stranding in HPSG. The former treat complement raising in the same lexicalist way as subject raising. The latter treat complement raising in the same way as complement extraction. Building on Van Eynde and Augustinus (2013) and Van Eynde and Augustinus (2014) we have argued that these reductions lead to problems and that complement raising had better be differentiated from both subject raising and complement extraction.

**Acknowledgment:** Many thanks are due to the editor and the anonymous reviewers of *Linguistics*. Preliminary versions of this paper were presented at the annual conference of the *Linguistic Society of Belgium* (Louvain-la-Neuve, May 13, 2016), the European Workshop on Head-driven Phrase Structure Grammar (Paris, March 24–25, 2017) and a seminar of the research group on Formal and Computational Linguistics (Leuven, June 27, 2017). I thank the participants for their comments and suggestions.

## Appendix

1. The queries to retrieve the circumpositions which surround their complement are of the following form:

(119) //node[@cat="pp" and node[@rel="hd" and @pt="vz" and number(@begin) < ../node[@rel="obj1"]/number(@begin)] and node[@rel="obj1" and number(@begin) < ../node[@rel="hdf" and @pt="vz"]/number(@begin)] and node[@rel="hdf" and @pt="vz"]]

The values of @rel are dependency relations, the values of @cat are phrasal categories, and the values of @pt are lexical categories.<sup>29</sup> < is the linear precedence relation. It is used to require the first adposition to precede the complement and to require the complement to precede the second adposition. The query was generated automatically by GrETEL on the basis of a relevant example. The query yields 270 hits for LASSY and 571 for CGN. 774 of these appear in Table 1. The remaining 67 can be divided in three groups. The largest one (50 tokens) concerns the combinations with *op ... na*, as used in (120).

(120) *Afrika was ... gekoloniseerd door Europese machten, op twee gebieden na*  
 Africa was ... colonised P European powers, P two  
 gebieden na  
 regions ADJ  
 ‘Africa was ... colonised by European powers, except for two regions’  
 (wr-p-e-i-0000051928.p.1.s.146)

We do not treat this as a circumpositional PP, since the meaning of *na* in this combination is not related to that of the adposition *na* ‘after’, as used in *na Pasen* ‘after Easter’. Instead it is related to that of the adjective *na* ‘near, close, dear’, as used in *alle kinderen zijn me even na* ‘all children are equally dear to me’. This adjective can be declined, as in *naë bloedverwanten* ‘close.DCL relatives’, and has comparative and superlative counterparts (*nader* ‘near.CMP’ and *naast* ‘near.SUP’). The combination *op ... na* is, hence, an AP which is headed by an adjective that selects a PP complement that is introduced by *op*. Other such adjectives are *belust* ‘keen’ and *gesteld* ‘keen’, as used in *op wraak belust* ‘keen on revenge’ and *op luxe gesteld* ‘keen on luxury’. A second group involves the combinations with *met ... mee* (5 tokens) and *op ... uit* (1 token), as used in (121–122).

<sup>29</sup> *vz* is short for *voorzetsel*, the Dutch term for adposition.

(121) *uw partner remigreert met u mee*  
 your partner remigrates P you P  
 ‘your partner remigrates together with you’  
 (dpc-svb-000431-nl-sen.p.27.s.1)

(122) ... *een burgemeester die op vrede uit is*  
 ... a mayor that P peace P is  
 ‘... a mayor that is eager for peace’  
 (fnf007134\_12)

They do not qualify as circumpositional PPs, since the order of the adpositions may be changed without any effect on well-formedness or meaning, as in (123–124).

(123) *uw partner remigreert mee met u*  
 your partner remigrates P you P  
 ‘your partner remigrates together with you’

(124) ... *een burgemeester die uit is op vrede*  
 ... a mayor that P is P peace  
 ‘... a mayor that is eager for peace’

The third group (11 tokens) concerns annotation errors and dysfluencies.

2. The queries to retrieve the stranded circumpositions take the form of regular expressions over strings, such as /naar toe/ and /naartoe/. To retrieve those with an incorporated R-pronoun we used more complex expressions, such as /\S{2,}naar toe/ and /\S{2,}naartoe/.<sup>30</sup> False hits were filtered out manually. They can be divided in three groups. The first consists of sequences in which the second adposition is the separable particle of a verb. In (125), for instance, *af* is part of the verb *afwijken* ‘diverge’.

(125) *de dialecten van de omringende plaatsen wijken er duidelijk*  
 the dialects P the surrounding places diverge it.R clearly  
*van af*  
 P SVP  
 ‘the dialects of the surrounding places clearly diverge from it’  
 (wiki-154.p.26.s.2)

<sup>30</sup> \S{2,} stands for a sequence of two or more non-white spaces. We choose for non-white spaces rather than for alpha-numeric characters in order to include combinations with *d’r*.

The second group consists of sequences in which the second adposition forms a unit with the words that follow it, rather than with the first adposition. This is, for instance, the case when it is a preposition followed by an NP, as in (126), or when it is a complementizer followed by a VP, as in (127).

- (126) *Pietje komt Jantje tegen [op de camping]*  
 Pietje comes Jantje SVP P the camping  
 ‘Pietje meets Jantje on the camping’  
 (fne000868\_130)

- (127) *alleen 's nachts kwam hij nog buiten [om naar de kroeg te gaan]*  
 only at night came he still SVP P the pub to  
 go  
 ‘he only came out at night in order to go the pub’  
 (wr-p-e-i-0000000332.p.4.s.52)

The third group consists of sequences in which the adpositions form a compound. This can be a prepositional compound, as in (128), or an adverbial compound, as in (129).

- (128) *vanaf 1882 gaat Ensor behoren tot de kunstkring “L’Essor”*  
 from.off 1982 goes Ensor belong to the art.circle “L’Essor”  
 ‘from 1882 (onward) Ensor becomes a member of the art group “L’Essor”’  
 (wiki-832.p.19.s.1)

- (129) *tussendoor vond hij nog de tijd om ...*  
 between.through found he still the time ...  
 ‘in between he found the time to ...’  
 (wr-p-p-i-0000000256.p.9.s.4)

Prepositional compounds invariably precede their complement. This differentiates them from the circumpositions which surround their complement if the complement is not an R-pronoun. Adverbial compounds do not take a complement.

## References

- Abeillé, Anne, Danièle Godard, Philippe Miller & Ivan A. Sag. 1998. French bounded dependencies. In Sergio Balari & Luca Dini (eds.), *Romance in HPSG*, 1–54. Stanford, CA: CSLI Publications.

- Augustinus, Liesbeth. 2015. *Complement raising and cluster formation in Dutch*. Leuven: Katholieke Universiteit Leuven dissertation.
- Augustinus, Liesbeth & Frank Van Eynde. 2014. Looking for cluster creepers in Dutch treebanks. *Computational Linguistics in the Netherlands Journal* 4. 149–170.
- Augustinus, Liesbeth, Vincent Vandeghinste & Frank Van Eynde. 2012. Example-based treebank querying. In *Proceedings of the 8th international conference on Language Resources and Evaluation (LREC)*, 3161–3167. May, 21–27, Istanbul.
- Bennis, Hans. 1986. *Gaps and dummies*. Tilburg: Katholieke Hogeschool Tilburg dissertation.
- Bouma, Gosse. 2000. Argument realization and Dutch R-pronouns. In Ronnie Cann, Claire Grover & Philip Miller (eds.), *Grammatical interfaces in HPSG*, 51–76. Stanford, CA: CSLI Publications.
- Bouma, Gosse & Gertjan van Noord. 1998. Word order constraints on verb clusters in German and Dutch. In Erhard Hinrichs (ed.), *Complex predicates in nonderivational syntax*. New York: Academic Press.
- Broekhuis, Hans. 2013. *Syntax of Dutch: Adpositions and adpositional phrases*. Amsterdam: Amsterdam University Press.
- Chomsky, Noam. 1970. Remarks on nominalization. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English transformational grammar*, 184–221. Waltham, MA: Ginn and Company.
- Engel, Ulrich. 1970. Regeln zur Wortstellung. In Ulrich Engel (ed.), *Forschungsberichte des Instituts für deutsche Sprache*, 1–148. Tübingen: Gunter Narr.
- Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative investigations*. Stanford, CA: CSLI Publications.
- Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij & Maarten C. van den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. Groningen and Deurne: Martinus Nijhoff and Wolters Plantyn.
- Hinrichs, Erhard & Tsuneko Nakazawa. 1989. Flipped out: AUX in German. In *Papers from the 25th Regional Meeting of the Chicago Linguistic Society*, 193–202. Chicago: Chicago University Press.
- Hinrichs, Erhard & Tsuneko Nakazawa. 1994. Linearizing AUXs in German verbal complexes. In John Nerbonne, Klaus Netter & Carl J. Pollard (eds.), *German in Head-driven phrase structure grammar*, 11–37. Stanford, CA: CSLI Publications.
- Hoekstra, Heleen, Michaël Moortgat, Bram Renmans, Machteld Schouppe, Ineke Schuurman & Ton van der Wouden. 2003. *CGN Syntactische annotatie*. Utrecht and Leuven.
- Höhle, Tilman. 1986. Der Begriff “Mittelfeld”, Anmerkungen über die Theorie der topologischen Felder. In Albrecht Schöne (ed.), *Kontroversen, alte und neue: Akten des 7. Internationalen Germanisten-Kongresses, Göttingen 1985*, vol. 3, 329–340. Tübingen: Niemeyer.
- Höhle, Tilman. 1997. Vorangestellte Verben und Komplementierer sind eine natürliche Klasse. In Christa Dürscheid, Karl Heinz Ramers & Monika Schwarz (eds.), *Sprache im Fokus: Festschrift für Heinz Vater zum 65. Geburtstag*, 107–120. Tübingen: Niemeyer.
- Jackendoff, Ray. 1977. *X-bar syntax: A study of phrase structure*. Cambridge, MA: MIT Press.
- Monachesi, Paola. 1998. Decomposing Italian clitics. In Sergio Balari & Luca Dini (eds.), *Romance in HPSG*, 305–357. Stanford, CA: CSLI Publications.
- Müller, Stefan. 1995. Scrambling in German. Extraction into the Mittelfeld. In *Proceedings of the 10th Pacific Asia conference on language, information and computation*, 79–83. Hong Kong: City University of Hong Kong.
- Müller, Stefan. 2002. *Complex predicates*. Stanford, CA: CSLI Publications.

- Oostdijk, Nelleke, Wim Goedertier, Frank Van Eynde, Louis Boves, Jean-Pierre Martens, Michaël Moortgat & Harald Baayen. 2002. Experiences from the Spoken Dutch Corpus Project. In M. Gonzalez Rodriguez & C.P. Saurez Araujo (eds.), *Proceedings of LREC 3*, 340–347. Paris: European Language Resources Association.
- Oostdijk, Nelleke, Martin Reynaert, Veronique Hoste & Ineke Schuurman. 2013. 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. Results by the STEVIN programme*, 219–247. Berlin and Heidelberg: Springer.
- Pollard, Carl J. & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Stanford, CA and Chicago: CSLI Publications and University of Chicago Press.
- Rentier, Gerrit. 1993. Dutch clitic pronouns and preposition stranding. In *Computational linguistics in the Netherlands. Papers from the Third CLIN Meeting*, 109–120. Tilburg: Institute for Language Technology and Artificial Intelligence, Tilburg University.
- Sag, Ivan A., Thomas Wasow & Emily Bender. 2003. *Syntactic theory. A formal introduction*, 2nd edn. Stanford, CA: CSLI Publications.
- Smeesaert, Hans, William Van Belle & Ingrid Van Canegem-Ardijns. 2014. Adpositional constructions of location and motion in Dutch. In Nicole Delbecq, Karen Lahousse & Willy Van Langendonck (eds.), *Non-nuclear cases*, 127–171. Amsterdam and Philadelphia: John Benjamins.
- Tseng, Jesse. 2005. Directionality and the complementation of Dutch adpositions. *Belgian Journal of Linguistics* 18. 167–194.
- Van Eynde, Frank. 1999. Major and minor pronouns in Dutch. In Gosse Bouma, Erhard Hinrichs, Geert-Jan Kruijff & Richard Oehrle (eds.), *Constraints and resources in natural language syntax and semantics*, 137–151. Stanford, CA: CSLI Publications.
- Van Eynde, Frank. 2004. Minor adpositions in Dutch. *Journal of Comparative Germanic Linguistics* 7. 1–58.
- Van Eynde, Frank. 2006. NP-internal agreement and the structure of the noun phrase. *Journal of Linguistics* 42. 139–186.
- Van Eynde, Frank. 2015. *Predicative constructions. From the Fregean to a Montagovian treatment*. Stanford, CA: CSLI Publications.
- Van Eynde, Frank. 2017. Adposition clusters in Dutch. In Hilke Reckman, Lisa L.S. Cheng, Maarten Hijzelendoorn & Rint Sybesma (eds.), *Crossroad semantics. Computation, experiment and grammar*, 77–92. Amsterdam and Philadelphia: John Benjamins.
- Van Eynde, Frank & Liesbeth Augustinus. 2013. Why and how to differentiate complement raising from subject raising in Dutch. In Stefan Müller (ed.), *Proceedings of HPSG 2013*, 222–242. Stanford, CA: CSLI Publications.
- Van Eynde, Frank & Liesbeth Augustinus. 2014. Complement raising, extraction and adposition stranding in Dutch. In Stefan Müller (ed.), *Proceedings of HPSG 2014*, 156–175. Stanford, CA: CSLI Publications.
- Van Noord, Gertjan, Gosse Bouma, Frank Van Eynde, Daniel de Kok, Jelmer van der Linde, Ineke Schuurman, Erik Tjong Kim Sang & Vincent Vandeghinste. 2013. Large scale syntactic annotation of written Dutch: Lassy. In Peter Spyns & Jan Odijk (eds.), *Essential speech and language technology for Dutch*, 147–164. Berlin and Heidelberg: Springer.
- Van Riemsdijk, Henk. 1978. *A case study in syntactic markedness*. Lisse: The Peter de Ridder Press.
- Zwarts, Joost. 1997. Complex prepositions and P-stranding in Dutch. *Linguistics* 35. 1091–1112.