Roots in Progress

Semi-lexicality in the Dutch and Afrikaans verbal domain

Published by

LOT
Kloveniersburgwal 48
1012 CX Amsterdam
The Netherlands

phone: +31 20 525 2461
e-mail: lot@uva.nl
http://www.lotschool.nl

Cover illustration by Regine Pots.

ISBN: 978-94-6093-338-7
NUR: 616

KU Leuven
Faculteit Letteren
Onderzoeksgroep Computationele en
Formele Taalkunde
CRISSP

Roots in Progress

Semi-lexicality in the Dutch and Afrikaans verbal domain

Cora Cavirani-Pots

Proefschrift tot het bekomen
van de graad van
Doctor in de Taalkunde

Promotor: Prof. Dr. Jeroen van Craenenbroeck

Leden: Dr. Liesbeth Augustinus
Prof. dr. Sjef Barbiers
Prof. dr. Theresa Biberauer
Prof. dr. Guido Vanden Wyngaerd
Prof. dr. Susanne Wurmbrand

The research reported here was supported by het Bijzonder Onderzoeksfonds of KU Leuven.

Voor opa en oma

Contents

List of Tables

List of Figures

List of Abbreviations

#	Number
γ	Gender
ϕ	Person
$\surd$	Root
3C	Third Construction
AUX	Auxiliary
And	Andative
ASP	Aspect
C	Complementiser
CA	Correspondence Analysis
D	Determiner
DEF	Definite
EP	Extraposition
EWG	European West Germanic
EXPL	Expletive
F	Functional *or* Feature
FVI	Functional Vocabulary Item
FE	Feature Economy
FIN	Finite verb
Fut	Future
HC	Hierarchical Clustering

IG	Input Generalisation
INF	Infinitival verb
INFL	Inflectional morpheme
IPP	Infinitivus Pro Participio
iT	Interpretable Tense feature
LVI	Lexical Vocabulary Item
MCA	Multiple Correspondence Analysis
MOD	Modal
NPI	Negative Polarity Item
Oblig	Obligatory
Opt	Optional
PASS	Passive
PC	Pseudocoordination
PL	Plural
PLD	Primary Linguistic Data
POS	Possessive
Prog	Progressive
PRT	Particle
PTCP	Participle
REFL	Reflexive pronoun
SG	Singular
T	Tense
uT	Uninterpretable Tense feature
UG	Universal Grammar
v	Verbalising head
V_1	Verb 1 of a verb cluster
V_2	Verb 2 of a verb cluster
V_3	Verb 3 of a verb cluster
V2	Verb Second
VR	Verb Raising

Acknowledgements

What a day, what a day!
- H. Pretorius

Doing a PhD is often metaphorically described as a journey, but I think I can literally say that my thesis has traveled. In the first three years in different forms at conferences throughout Europe, but it made its ‘grand tour’ during my final year, from Leuven to Cambridge, South Africa, Sardegna, some pit stops in Ortonovo and Leuven, back to Sardegna, and finally back to Leuven again. Speaking of this grand tour directly leads me to three people I would like to thank here first, as they made it possible for me to do this PhD the way I’ve done it, and furthermore made the process smoother, more interesting and more fun than it would have been without them.

The first of these three people is my supervisor Jeroen van Craenenbroeck, who really deserves the first and biggest thanks of all. Jeroen, thank you for your endless guidance and support throughout these four years, for expressing your high expectations and (fortunately) also your trust in me as a PhD student, for always sharing my enthusiasm about some new topic or phenomenon that attracted my attention, and for always pushing me to make my ideas more concrete and precise. You let me follow my own interests and passions, allowed me to change and twist the initial project outline accordingly, accepted (or had to accept) my stubbornness, and allowed me not to be around for the biggest part of this last year in which I’ve visited (both personally and linguistically speaking) significant places. I’m especially very grateful for the very close reading of the drafts of my chapters, which you’ve continued to do until the very end. I’ve really learned so much from you, which I will always carry with me in whatever direction the future brings me. And I shouldn’t forget: also thanks for letting Edo and me squat your guest room and part of your garden with our furniture, and for running around at our wedding to get a beautiful portrait of all our guests.

Then, Theresa. Three and a half years ago, I enrolled in a research discussion seminar at the LOT school with Theresa Biberauer, who was by then for me only a 'big name' in diachronic syntax, whose papers I had read during my MA, but had never met in real life. Well, Theresa is super amazing in real life, I can tell you that. When I presented my research topic, she told me: 'Ah, where you (read Dutch) have *te*, we (read Afrikaans) have *en*, maybe you should do a comparative study on this!'. This comment was the seed that now has grown out to a substantial part of this thesis. One year later, Theresa came to Brussels to give a lecture series at our campus, which I then felt and still feel was the most inspiring week of my PhD. Her enthusiasm, her deep insight on abstract and big picture topics, her great knowledge of a massive variety of data, and her endless energy both regarding linguistics and life in general, are a great source of inspiration for me. Again a bit more than a year later, she made it possible for me to come to Cambridge for two months, and furthermore was willing to share her house with me and Edo, which has been such a wonderful experience. Here I would also like to thank André for allowing two extra linguists in the house, for the fun discussions on – among other things – squirrel rights and the glorious taste of vegetarian food (glad we agree on that), for filling in that endless questionnaire while growing more and more concerned about Thomas' unhealthy relationship with his chin (don't worry, it turned out he had a mosquito bite), and for secretly accepting and maybe even admiring our Free-style approach to cooking. Also thanks to dear Danieltjie, for distracting us from linguistics by running around the house like crazy, for making us familiar with great hits such as 'Baby shark (tutududuuutudu)' and 'Shot gun', for needing a partner to decorate the Christmas tree, and just for being the cute little Roman soldier that he is. Back to Theresa: I would like to thank you for all these wonderful experiences, all our inspiring discussions over the last four years, and especially for supporting me throughout the last phase of writing with your usual enthusiasm, feedback and positive vibes. Baie baie dankie!

The third main figure that made it possible for this thesis to be finished is my husband Edo. Thank you for following my progress every step of the way, for coming with me to Cambridge and to South Africa, and for taking me to Sardegna, where the delicious food was the perfect remedy when I was stressed about writing up the thesis. Also thank you for our long walks in spring, where you listened to me rambling on about *te*, *hoeven* and *zitten*, patiently providing suggestions while picking wild asparagi for the frittata you were going to prepare for lunch (or the pasta, or the risotto, or the salad, so many asparagi!). In our stay there during the summer, thank you for our discussions on roots and features, on grammaticalisation and language change, and for calming me down (by suggesting me to take a sea ride on the *tucano*) when the writing pressure and the failing anxiety became too much for me to handle. Thank you especially for these last few months of my very intense writing, where you fed me, took so many things out of my hands for which I didn't have the time or energy anymore, and for always cheering me up when

I was in a pessimistic or negative mood. I know that if I would say: ‘I couldn’t have done it without you’ you would have told me that’s not true, so let me say at least that these last years of my PhD would definitely not have been so much fun if you hadn’t been there.

Let me briefly return some of the destinations of the ‘tour’ my thesis. First destination (well, not diachronically, but in terms of significance): South Africa, and more specifically, Pinelands, where we were hosted by two of my favourite people, Erin and Helgard. Spending more than a month together has been such an amazing experience, which brought together so many good things: climbing, eating amazing food, learning how to prepare dinner during load shedding when the kitchen has electric cooking plates, learning to braai, drinking good wine, watching the most beautiful sky full of stars in Cederberg, being woken up by Helgard bringing freshly made coffee, sharing good stories and cheering each other up when needed. Baie dankie vir alles! Also, dear Erintjie, thank you for all the long-distance support during my writing period, which really has meant a lot to me. It is always so nice when someone exactly understands how you feel. Speaking of our stay in South Africa, I would also like to thank Jac and Anneline for their hospitality, and for the nice chats on linguistic and personal topics. The discovery of me accidentally ending up with exactly the same research topic as one of my father’s oldest friends remains one of my favourite stories related to my PhD.

Regarding destination Sardegna: thanks to Domenico, Marinella and Dino for making us feel part of the locals in Giuncana, and specifically to Domenico for the adventurous hikes on Monte Ruju. Also concerning Italy, destination Ortonovo: many many thanks to Oriana and Agostino, and all other Cavirani’s, for accepting me always having to ‘study’ while being there, and for always asking me how it was going.

Returning to destination Cambridge, I want to thank all the people who made the time to discuss my research with me, which has been of great help in developing my analysis: Ad Neeleman, Daniel Harbour, David Hall, Coppe van Urk, David Adger, Hagit Borer and Jenny Cheshire. Also thanks to Kim, Libero, Oliver, Gigi, Curtis and Lovisa for the nice moments together.

Then, the starting and endpoint: Leuven and surroundings. I want to thank my office mate and friend Ngum for four years of laughter, good advice, and long chats on everything except linguistics. I will never forget. Also a huge thanks to all (former and current) CRISSP members for making me feel at home at KU Leuven, for all the linguistic input, and all the bowling, escape-rooming, laser-shooting, hiking and many lunches and especially dinners: Dany, Guido, Tanja, Jeroen, Hans, Jolijn, Will, and Edo. An extra thanks to Jolijn for continuing to send me supporting messages during the last stage of my writing. Also a big thanks to Liesbeth for our very pleasant cooperation and fun discussions about the similarities and differences between Dutch and Afrikaans. Infinitive thanks to Lili, who supported me during these four years on an almost daily basis, and with whom I’ve done so many fun things for distraction. Your cheering me up when I really didn’t feel like writing anymore helped me so much to push through until the end, thank you so much for that. Of course, your great compositions (my favourite still being 'Ry-ry-ry-ryan air') made things much easier as well. Concerning Leuven I also want to thank the members our little Emigrati family for all the great dinners and mental support: Giulia, Antonella, Grace, Luciano and Steve, grazie mille per tutto! Slightly further away, though still in Flanders: Kitty, thank you for all your checking in and good advice. Somehow I always feel relieved from all my worries when I've spend some time with you. Specifically related to the four years in which I've worked at the Erasmushuis, I want to thank Sylvia and Sonja for everything they've done for me.

Even though I haven't been writing my thesis in the Netherlands, the Basque Country or Austria, these locations are connected to a number of people I want to thank here too. Astrid, mama, Harry and Regine, thank you for all the support during these years of my PhD. Harry, zoals je kunt zien heb ik dan eindelijk 'mijn werkstuk' af. Also, back in the Netherlands, I want to thank Karlijn for regularly checking in on me regarding my thesis throughout all four years, and for always sending me supporting messages. In the Basque Country, another two of my favourite people live, who supported me by sharing bread, going on climbing adventures, and by their round videos: thank you dear Amaia and Varun. Turning to Austria: Regine, besides the support, also a big thanks for the beautiful cover photo, and for annotating the Afrikaans data for me.

Speaking of data, I am very grateful for all the Dutch and Afrikaans speakers who filled in my questionnaires. Regarding the Dutch questionnaires, I want to thank the editors of *Taalpost* for distributing the first questionnaire in their newsletter, and Maarten van der Peet for building and distributing the second one. I also want to thank Benito, Erin and Kitty for annotating data.

Finally, I want to thank a number of linguists who during these four years have been advising me or have taken the time to discuss my ideas with me. I would like to thank Marjo and Benedikt for being in my advisory committee. I am grateful to Susi Wurmbrand, Paola Merlo, Bronwyn Bjorkman, Charles Yang, Norbert Corver, Martin Salzmann, Lotte Dros-Hendriks, Marijke de Belder, Heidi Klockmann, Gillian Ramchand, and again, Marjo for the useful and interesting discussions regarding different aspects of my thesis.

Leuven, 19 December 2019

CHAPTER 1

Introduction

1.1 The concept of semi-lexicality

This thesis investigates semi-lexicality in the Dutch and Afrikaans verbal domain. A Dutch example of a semi-lexicality is given in (1).

(1) Ik heb de hele dag ***zitten*** te **lezen**.
I have the entire day sit to read
'I have been reading the entire day.'

As can be seen in the example, the posture verb *zitten* 'sit' is used here to indicate progressive or durative aspect of the lexical verb of the clause: *lezen* 'read'. The subject of the sentence has been reading the entire day. The fact that the verb *zitten* expresses aspectual information in (1) shows that the verb is used functionally rather than lexically in the sentence. However, *zitten* is not completely functional in sentences like the one in (1). This becomes clear in a sentence like (2).

(2) ??Ik heb de hele dag ***zitten*** te **zwemmen**.
I have the entire day sit to swim
'I have been swimming the entire day.'

The semantics of the lexical verb of the clause, *zwemmen* 'swim', are incompatible with a seated position. Many Dutch speakers do not allow the verb *zitten* to indicate progressive or durative aspect when the semantics of the lexical verb are incompatible with the lexical meaning of *zitten* (Lemmens 2005, Haeseryn et al. 1997). The fact that many Dutch speakers find sentences like (2) very odd shows that the semantics of *zitten* are still partly retained when it is used to express progressive or durative aspect. Thus, in sentences like (1)–(2), *zitten* behaves like a functional verb in that it signals the type of aspect of the lexical verb, while at the same time behaving like a lexical verb in expressing its own lexical semantics. Vocabulary items which show both functional and lexical properties are called *semi-lexical* (Klockmann 2017). Thus, Dutch *zitten* can be seen as a semi-lexically used verb. Note, though, that *zitten* can also be used completely lexically (3).

(3) Ik heb de hele dag op de bank ***gezeten***.
I have the entire day on the couch sit.PTCP
'I have been sitting on the couch the entire day.'

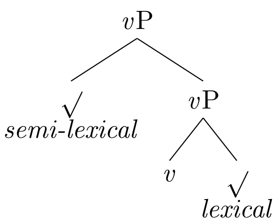
In this sentence, *zitten* is the lexical verb of the clause: the subject has been sitting the entire day. In this thesis, I henceforth make a distinction between the *lexical use* – as in (3) – and the *semi-lexical use* – as in (1) – of a given vocabulary item.

1.2 The theoretical proposal in a nutshell

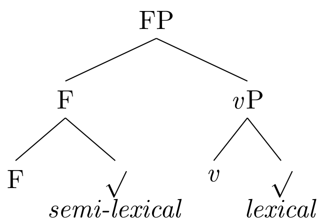
In this thesis, I present a theoretical analysis of semi-lexically used verbs in Dutch and Afrikaans. I thereby adopt the tenet of the Distributed Morphology framework and the Exo-Skeletal Model that lexical vocabulary items are the spell out of roots, and functional vocabulary items the spell out of syntactic features on functional heads (Halle and Marantz 1993, Harley and Noyer 1999, Borer 2005a). As we have seen in the previous section, the semi-lexical use of a vocabulary item shows both functional and lexical properties. The question is thus how semi-lexicality should be analysed in terms of roots and syntactic features. This thesis keeps the roots versus syntactic features division between lexical and functional vocabulary items, but proposes that under certain circumstances lexical vocabulary items (i.e. roots) can be Merged in the functional domain of another lexical item (cf. Klockmann (2017)). When this happens, this root is used semi-lexically.

A central claim in my theoretical proposal is that there are two stages of semi-lexicality. Both stages have different underlying syntactic structures. Furthermore, the first stage of semi-lexicality is the diachronic ancestor of the second one. This means I analyse semi-lexicality as the result of grammaticalisation (following among others Haider (2001), Hagemeijer (2001), Klockmann (2017)), in which the two stages of semi-lexicality should be seen as two (very early) consecutive steps along a grammaticalisation path. In the first stage, a root is Merged very low in the functional domain of the main lexical root of the clause. In the second stage, the semi-lexically used root is merged with a functional head in a separate workspace, forming a complex head (cf. Song (2019)). This complex head is then Merged in the functional domain of the main lexical root of the clause. The abstract structures of the two stages of semi-lexicality (applied to the verbal domain, which is the main focus of this thesis) are illustrated in (4) and (5).

(4) **Semi-lexical stage I**



(5) **Semi-lexical stage II**



1.3 The empirical scope of the thesis

The empirical scope of this thesis is the verbal domain of Dutch and Afrikaans. Specifically, I look at semi-lexical restructuring in these two languages. Roughly speaking, restructuring is a phenomenon in which two verbs share one verbal extended projection (cf. Wurmbrand (2001, 2004) – the phenomenon of restructuring is discussed in more detail in Chapter 2). I investigate the extent to which semi-lexical restructuring leads to morphosyntactic variation and optionality in both languages.

In Dutch, I investigate the semi-lexical use of two verbs: *hoeven* 'need' and *zitten* 'sit' (cf. (1) above). In their semi-lexical use, these verbs select an infinitive introduced by *te* 'to', illustrated for *hoeven* in (6).

(6) Hij zal morgen niet ***hoeven te*** **gaan** voetballen.
He will tomorrow not need to go play.football
'He won't need to go play football tomorrow.'

Based on the data of two questionnaire studies, I show that there is a high degree of optionality among and between speakers regarding the presence and position of *te*. For example, many speakers allow *te* to remain absent (7-a), or to occur on *hoeven* itself rather than on the infinitive following *hoeven* (where it should occur based on selectional requirements) (7-b).

(7) a. Hij zal morgen niet ***hoeven*** **gaan** voetballen.
He will tomorrow not need go play.football
'He won't need to go play football tomorrow.'
b. Hij zal morgen niet ***te hoeven*** **gaan** voetballen.
He will tomorrow not to need go play.football
'He won't need to go play football tomorrow.'

The same patterns are found with *zitten* (8-a)–(8-b).

(8) a.

Hij	zal	morgen	lang	op	de	bus	moeten	***zitten***	**wachten**.
He	will	tomorrow	long	on	the	bus	must	sit	wait

'He will have to wait for the bus for a long time tomorrow.'

b.

Hij	zal	morgen	lang	op	de	bus	moeten	***te zitten***	**wachten**.
He	will	tomorrow	long	on	the	bus	must	to sit	wait

'He will have to wait for the bus for a long time tomorrow.'

I investigate the relation between the degree of semi-lexicality of these two verbs (i.e. whether they are in the first or second stage, and whether they are transitioning from one stage to the other or not) and the degree of variation and optionality in their morphosyntax.

I do the same for pseudocoordination constructions in Afrikaans with the motion verb *loop* 'walk' and the posture verbs *sit* 'sit', *staan* 'stand' and *lê* 'lie'. The motion verb and the posture verbs in these constructions are semi-lexical, in the sense that they indicate a type of aspect of the lexical verb, but at the same time still retain (part of) their own semantics. Based on a corpus study and a questionnaire study, I show that there is a high degree of morphosyntactic variation and optionality in these constructions. Two examples of this are the presence vs. absence of the coordinator *en* 'and', and the presence vs. absence of the perfect participle marker *ge-*. An example involving the motion verb *loop* is given in (9).

(9)

Ek	het	gister	baie	***(ge-)loop***	**(*en*)**	**praat**.
I	have	yesterday	a.lot	GE-walk	and	talk

'I've been (walking and) talking a lot yesterday.'

As was the case for Dutch, I investigate the role of the degree of semi-lexicality of these verbs in determining the amount of variation and optionality in the concomitant constructions.

1.4 The outline of the thesis

The **theoretical proposal** of this thesis is presented in **Chapter 2**. In this chapter, I discuss two recent approaches to semi-lexicality in which the distinction between roots and syntactic features is implemented, namely De Belder (2011) and Klockmann (2017). I furthermore introduce the phenomenon of restructuring, and specifically Wurmbrand (2001)'s three-way division between lexical, semi-lexical and functional restructuring, which I adopt in this thesis. I show that Dutch motion and posture verbs (like semi-lexical *zitten* in (1)) pose a problem for her analysis of semi-lexical restructuring. The last section of the chapter is devoted to the main proposal of this thesis, namely that we need to distinguish two stages of semi-lexicality, which have different underlying syntactic structures.

The theoretical proposal is put to work in **two case studies**. **Case study I (Chapters 3–6)** deals with semi-lexicality in the Dutch verbal domain. More specifically, I investigate how the different degrees of semi-lexicality of Dutch *hoeven* ‘need’ and *zitten* ‘sit’ have a different impact on the variation and optionality in the presence and placement of the infinitival marker *te* in their verb cluster. **Case study II (Chapters 7–11)** looks at semi-lexicality in the Afrikaans verbal domain. The main focus of this case study lies in the impact of the degree of semi-lexicality of the motion verb *loop* ‘walk’ and the posture verbs *sit* ‘sit’, *staan* ‘stand’ and *lê* ‘lie’ on the morphosyntactic variation and optionality in pseudocoordination constructions. The three variation and optionality patterns that are investigated are the presence vs. absence of *en* ‘and’, the presence vs. absence of the perfect participle marker *ge-*, and the type of Verb Second the construction undergoes (quirky vs. normal V2).

The thesis concludes with a **general discussion** in **Chapter 12**, in which I focus on the theoretical consequences of my proposal for the theory of grammaticalisation within the Generative framework. Furthermore, I discuss the different types of morphosyntactic optionality that the two case studies have brought to our attention: some types of optionality are due to ongoing grammaticalisation, others to spell-out mechanisms, and others still to other factors. I end the chapter with a number of possible directions for future research.

CHAPTER 2

Theoretical proposal

2.1 Introduction

In this chapter, I present the theoretical proposal of this thesis, namely that there are two stages of semi-lexicality, which have different underlying syntactic structures. In section 2.2, I present two recent accounts of semi-lexicality: De Belder (2011) and Klockmann (2017). Both of them define semi-lexicality within the broader context of the distinction between lexical words as (the spell-out of) roots and functional words as (the spell-out of) syntactic features. In section 2.3 I turn to semi-lexicality in the verbal domain, specifically to the phenomenon of restructuring. I first introduce the phenomenon by discussing Wurmbrand (2001)'s analysis of semi-lexical restructuring. I then show that this analysis cannot adequately deal with semi-lexical restructuring verbs in Dutch. In section 2.4 I present my own proposal regarding semi-lexical restructuring.

2.2 Two recent accounts of semi-lexicality

In any approach of semi-lexicality that adopts a division between functional words and lexical words, the pressing question is how to syntactically analyse semi-lexical words, given that they have properties of both functional and lexical words. In this section I discuss two recent approaches that address this question. More specifically, in subsection 2.2.1, I discuss the approach of semi-lexicality of De Belder (2011), according to which semi-lexicality is the result of functional words being used in the position of a lexical word. In subsection 2.2.2, I discuss the approach of Klockmann (2017), in which semi-lexicality is the result of a root being specified with a (number of) syntactic feature(s) in the lexicon.

2.2.1 De Belder (2011)

Building on Borer (2005a,b)'s Exo-Skeletal Modal, De Belder (2011) implements the general lexical/functional divide as follows. In narrow syntax, there are functional head positions and root positions. Functional words (for example the English determiner *the*) are called functional vocabulary items (henceforth FVIs), and they spell out syntactic features on functional heads. Lexical words (for example the English noun *book*) are called lexical vocabulary items (henceforth LVIs) and they spell out of roots in root positions. Abstractly, a single cycle in a syntactic derivation in narrow syntax exists of one or more functional heads and a root position. In a derivation like the one in (1), in which the functional head is specified for a [+def]-feature, the FVI *the* can be the spell out the [+def]-feature on the functional head, and the LVI *book* the root in the root position. The result is the nominal constituent *the book*.[1]

(1)

$$[_{FP_{[+def]}}\ F_{[+def]}\ \ \sqrt{}\ ]$$

(2)

$$[_{FP_{[+def]}}\ F_{[+def]}{\Rightarrow}\textit{the}\ \ \sqrt{}{\Rightarrow}\textit{book}\ ]$$

De Belder (2011) proposes that not only LVIs but also FVIs can be used in root positions.[2] She gives the following Dutch examples as empirical evidence (see also De Belder and Van Craenenbroeck (2015) for more examples and for examples from languages other than Dutch).

(3) Ik heb het **waarom** van de zaak nooit begrepen.
I have the why of the case never understood
'I have never understood the motivation behind the case.'

(4) Martha is mijn tweede **ik**.
Martha is my second I
'Martha is my best friend.'

(5) De studenten **jij**-en onderling.
the students you-INF mutually
'The students are on a first-name basis with one another.' (De Belder 2011:42)

[1] Note that I abstract away here from the technical details of De Belder (2011)'s work. The interested reader is referred to De Belder (2011) and De Belder and Van Craenenbroeck (2015) for in-depth argumentation and illustration.

[2] Again, for the technical argumentation that this is possible, the reader is referred to De Belder (2011) and De Belder and Van Craenenbroeck (2015).

In (3), a *wh*-word is used as a noun, i.e. the FVI *waarom* is merged in a root position under a nominal structure, and similarly for the personal pronoun *ik I* in (4). In (5), a personal pronoun is merged in a root position under a verbal structure, resulting in *jij-en* as a verb that refers to the event of students being on a first-name basis with each other.

De Belder (2011) argues that semi-lexicality is the result of a FVI being used in a root position.[3] One of the illustrations of semi-lexicality she discusses is the Dutch word *heel* 'whole' (see also Zwarts (1992), Den Dikken (2002)), which can be used both as a quantifier (i.e. its functional use) and as an adjective (i.e. its lexical use). The former is illustrated in (6) and the latter in (7).

(6) Ik heb **heel** het huis gepoetst.
I have whole the house cleaned
'I have cleaned the entire house.'

(7) Het **heel-e** bord is veel waard, het kapot-e bord niet.
the whole-INFL plate is much worth the broken-INFL plate not
'The intact plate is worth a lot, the broken one isn't. (De Belder and Van Craenenbroeck 2014:187)

The lexical adjective *heel* in (7) can be recognised as being an adjective by the fact that it shows inflection (*-e*), like the other adjective in the sentence *kapot-e* 'broken'. As can be seen from (6), the quantifier *heel* occurs to the left of the determiner, and can thus be assumed to occupy a position high in the functional sequence of the noun (*huis* 'house' in this example). Other universal quantifiers, like *al* 'all', can also occur to the left of the determiner (8), suggesting that *heel* is indeed a quantifier in the example in (6). Furthermore, the quantifiers *heel* and *al* are in complementary distribution, with the former only occurring with definite singulars (9-a), and the latter with definite plurals and definite mass nouns (9-b)–(9-c).

(8) **Al** de chocolade.
all the chocolate
'All the chocolate.'

(9) a. ***Al** / **heel** de regio.
All whole the.DEF region.SG
'The entire region.'
b. **Al** / ***heel** de regio's.
All whole the.DEF regions.PL
'All the regions.'
c. **Al** / ***heel** de chocolade.
All whole the.DEF chocolate.MASS
'All the chocolate.' (De Belder and Van Craenenbroeck 2014:189-190)

[3] See Acedo-Matellán and Real-Puigdollers (2019) for the opposite proposal, in which semi-lexicality is the result of LVIs being used in a functional head position.

Summing up, Dutch *heel* is a quantifier when used to the left of the determiner in a noun phrase. However, it can also be used as an adjective, in which case it can be modified by a degree modifier (10), just like other adjectives can (11).[4]

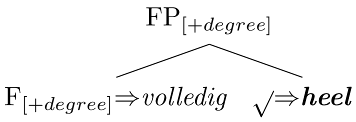
(10) Het bord was nog ***volledig*** **heel**.
the plate was still completely whole
'The plate was still completely intact.'

(11) Het bord was nog ***volledig*** **leeg**.
the plate was still completely empty
'The plate was still completely empty.' (De Belder and Van Craenenbroeck 2014:187)

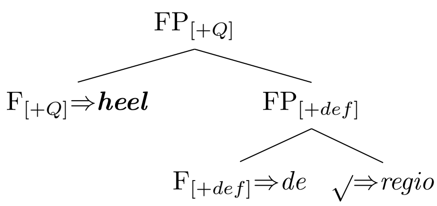
All the examples discussed so far combined thus show that *heel* can be used as a FVI (a quantifier) and as a LVI (an adjective). Important to note here is that De Belder (2011)'s definition of semi-lexicality is not equal to the one I have set in chapter 1. For her, semi-lexicality means a vocabulary item being completely functional in one context (i.e. when it is Merged in a functional head position), and being completely lexical in another (i.e. when it is Merged in a root position). In contrast, I define semi-lexicality as the semi-lexical use of a lexical vocabulary item, which results in behaviour which is both functional and lexical at the same time.

De Belder (2011)'s analysis of the semi-lexicality of *heel* is as follows: it is a FVI that in its adjectival use has been inserted into a root position under an adjectival functional projection. When it is used as a quantifier, it is inserted into a functional head of quantification, spelling out the [Q]-feature on that functional head. The former use is illustrated in (12) for the adjectival phrase *volledig heel* 'completely intact', the latter in (13) for the noun phrase *heel de regio* 'the entire region'.

(12)



(13)



De Belder (2011) can thus account for the dual, i.e. semi-lexical, behaviour of vocabulary items such as Dutch *heel*: vocabulary items that can both be used functionally and lexically are FVIs. They are used functionally when inserted in a functional head position, and lexically when inserted in a root position.

[4]See De Belder (2011) and De Belder and Van Craenenbroeck (2014) for more evidence that in this use of *heel*, it has the status of an adjective.

However, there are also vocabulary items whose functional use is not identical to the functional use of 'real' FVIs. Instead, they show mixed properties and behave syntactically in between FVIs and LVIs. An English example, which is extensively discussed in Klockmann (2017) is the English word *bunch* (see also Riemsdijk (1998a), Stavrou (2003), Alexiadou et al. (2008)). This element can be used both lexically, as a noun (14), and functionally, as a quantifier (15).

(14) The flowers were arranged in a beautiful **bunch**.

(15) A **bunch** of chickens ran down the mountain. (Klockmann 2017:4)

In (14), *bunch* clearly is a noun, as it is modified by the adjective *beautiful*. In (15), *bunch* is used as a quantifier, indicating a quantity of the noun *chickens*, namely something like a small group.[5]

Even though *bunch* can be used to indicate the quantity of a noun, in its quantificational use it is clearly different from real quantifiers, such as *many*. As is shown in (16), *bunch* requires an indefinite determiner in its quantificational use, as well as the insertion of *of* between itself and the noun it modifies. Both of these properties are obligatory. In contrast, real quantifiers, like *many*, show the opposite behaviour, with both the determiner and *of* being ungrammatical in combination with *many* (17).

(16) a. A **bunch** of chickens.
b. ***Bunch** chickens.
c. *A **bunch** chickens.
d. ***Bunch** of chickens.

(17) a. **Many** chickens.
b. *A **many** of chickens.
c. *A **many** chickens.
d. ***Many** of chickens. (Klockmann 2017:216)

The fact that the quantificational use of *bunch* does not show the same morphosyntactic behaviour as real quantifiers like *many* is unexpected in De Belder (2011)'s account of semi-lexicality. Recall that in her analysis, semi-lexicality is the result of a single FVI sometimes being used to spell out a functional head position (its functional use), and sometimes being used to spell out a root position (its lexical use). Under such an account, we would thus analyse English *bunch* as a FVI with a quantificational ([Q]-)feature. This means that it should be perfectly possible for the FVI *bunch* to be inserted in a functional head position specified for a [Q]-feature, and thus to occur in exactly the same structural configuration as quantifiers like *many* (as was the case for *heel* and *al* in (6) and (8) above). For English *bunch* it is clearly not the case that it can occur in the same structural position as the quantifier *many*, given that *many* never takes an indefinite determiner, and never combines with *of*.

[5] See Klockmann (2017:chapter 6) for detailed discussion of the semi-lexical quantificational use of the English nouns *bunch*, *lot* and *number*.

Note that fully lexical nouns also require the presence of an indefinite determiner and *of* when followed by another noun, as illustrated in (18) for the noun *study*.

(18) A **study** of possessives is worth pursuing. (Klockmann 2017:4)

On the surface, this structure looks very similar to one in which *bunch* indicates the quantity of a noun, like in (15), repeated here in (19).

(19) A **bunch** of chickens ran down the mountain. (Klockmann 2017:4)

However, unlike the noun *study* (20), *bunch* cannot occur in the plural form (21).

(20) **Studies** of possessives are worth pursuing.

(21) ***Bunches** of chickens ran down the mountain.

Furthermore, while *study* can control agreement (22), *bunch* cannot (23). That is, in (22), the finite verb Agrees with the singular noun *study*, and not with the plural noun *possessives* in its complement. In contrast, in (23), the finite verb agrees with the plural noun *chickens*, and not with *bunch*, which appears in its singular form.

(22) A **study** of possessives was/*were worth pursuing.

(23) A **bunch** of chickens *was/were found on the trail. (Klockmann 2017:4, 217)

Thus, the quantificational use of *bunch* has a distinct morphosyntax from lexical nouns that select another noun: it is necessarily singular, and it cannot control agreement (Klockmann 2017:234). The morphosyntactic behaviour of this semi-lexical use of *bunch* is thus in between that of the lexical use of *study* on the one hand, and that of fully functional quantifiers like *many* on the other. It is precisely this combination of functional and lexical properties *within one and the same example*, which I have taken to be a hallmark of semi-lexicality (see chapter 1), that is hard to capture under De Belder (2011)'s account.

Another property of *bunch* that is unexpected in De Belder (2011)'s approach is that in its quantificational use, *bunch* still partly retains its lexical semantics. That is, it can only be used to signal a relatively large quantity of things that form a collection together, corresponding to its lexical meaning of an actual bunch (Klockmann 2017:234). For example, an actual bunch of flowers can contain many, but not a huge amount of flowers, as the bouquet should still be able to remain together and fit into the hands of a person. The presence of lexical semantics in the functional use of a vocabulary item is unexpected in the approach of De Belder (2011), because in this approach vocabulary items that can be both used functionally and lexically are always FVIs, and FVIs do not have lexical semantics.

To account for the semi-lexicality of English *bunch*, we need an approach that can deal with vocabulary items whose morphosyntactic behaviour when used functionally is distinct both from that of real FVIs used in functional head positions and from that of LVIs used in root positions. The approach of De Belder (2011) is not able to do that. Note that there are clear parallels between the semi-lexical use of *bunch* and that of the Dutch and Afrikaans verbs investigated in this thesis. For example, as I have discussed in Chapter 1, Dutch *zitten* 'sit' and the Afrikaans motion and posture verbs can be used to indicate the aspect of the lexical verb, but still retain at least part of their own lexical semantics. In the two case studies of this thesis, I also show that the semi-lexical use of these Dutch and Afrikaans verbs share (morpho)syntactic behaviour both with FVIs of the verbal domain (auxiliaries and modals) and with LVIs (lexical verbs). The fact that Dutch and Afrikaans verbs partly retain their semantics when used semi-lexically, and that their (morpho)syntactic behaviour is in between that of FVIs and LVIs makes them incompatible with an approach to semi-lexicality such as the one of De Belder (2011).

An account that offers an analysis of the semi-lexicality of *bunch*, and that thus might be a better fit for the Dutch and Afrikaans verbal domain as well, is that of Klockmann (2017). In the next subsection, I summarise the main points of her analysis, indicate which parts of her analysis I will adopt in my own, and which parts are in need of refinement or require a different explanation.

2.2.2 Klockmann (2017)

Klockmann (2017) uses the same distinction as De Belder (2011) between roots and syntactic features to separate LVIs from FVIs. In contrast to De Belder (2011), though, who analyses semi-lexicality as the use of a FVIs in a root position, Klockmann (2017) takes semi-lexicality to be defined in the lexicon. That is, there are semi-lexical vocabulary items, besides FVIs and LVIs. Semi-lexical vocabulary items are seen as the exception to the lexical-functional dichotomy in its most literal sense: they are roots that are specified in the lexicon for a syntactic feature or feature bundle (Klockmann 2017:25). Those syntactic features can be either positive or negative (Klockmann 2017:42).

As an illustration of the proposal, let us return to the semi-lexical use of *bunch*, and compare its lexical entry to that of a LVI, such as *study*, and to that of a FVI, such as *many*. The lexical entries of these three vocabulary items are given in Table 2.1.[6]

The lexical noun *study* is a LVI; it spells out a featureless root. The quantifier *many* is a FVI, and it is the spell-out of the quantificational feature [Q]. The word *bunch* (in its semi-lexical use) is taken to be a semi-lexical vocabulary item, specified in the lexicon as a root with a [Q]-feature.[7] In the previous

[6]The feature specification of *bunch* consists of more than the one syntactic feature indicated here in Klockmann (2017). See for this more complex specification. For the sake of illustrating the difference between lexical, semi-lexical and functional vocabulary items under Klockmann (2017)'s approach, though, the simplified representation in Table 2.1 suffices.

[7]Recall that *bunch* can also be used lexically, which is unexpected if it spells out a root

lexical	semi-lexical	functional
study	*bunch*	*many*
$\surd$	$\surd$ [Q]	[Q]

Table 2.1: Part of the English lexicon in Klockmann (2017)'s approach

subsection we have seen that *bunch* behaves differently morphosyntactically from both LVIs and FVIs. Before illustrating Klockmann (2017)'s analysis of the 'in between' status of *bunch*, I first need to discuss a number of theoretical assumptions she makes related to the structure of nominal constituents in English. She assumes an adapted version of the Universal Spine of Wiltschko (2014), specifically the one proposed by Hachem (2015) for nominal phrases.[8] This Universal Spine is given in (24) (Klockmann 2017:40).

(24) **Universal Spine for nominal phrases**

DP = ANCHORING domain
D QP = QUANTITY domain
Q ϕP = CLASSIFICATION domain
ϕ $\surd$ = IDENTIFICATION domain

The universal structure of a nominal phrase exists of a root (in the identification domain), phi-information (like number and gender, in the classification domain), quantity information (like *many*, in the quantity domain), and information related to definiteness (like indefinite or definite determiners, in the anchoring domain).[9] There are a number of assumptions that Klockmann (2017) makes regarding this spine that are important for her analysis of semi-lexicality. First, not all domains are necessarily present in each nominal phrase: a nominal phrase does not have to include quantity information (e.g. *the books*), or complete phi-information (e.g. mass nouns are often analysed as structures lacking a number layer). Second, a domain can consist of different projections in different languages. For example, a language can use two projections for

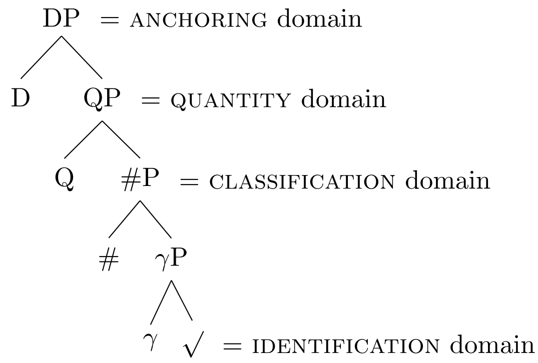
specified with a syntactic feature. See Klockmann (2017:161, fn. 2) for discussion. She proposes that it is possible to ignore a syntactic feature as long as a vocabulary item is still a root.

[8] Hachem (2015) relabels Wiltschko (2014)'s original 'point-of-view' domain as the quantity domain, and locates phi-information (number and gender) in the classification domain, rather than in 'point-of-view', where it is located in Wiltschko (2014).

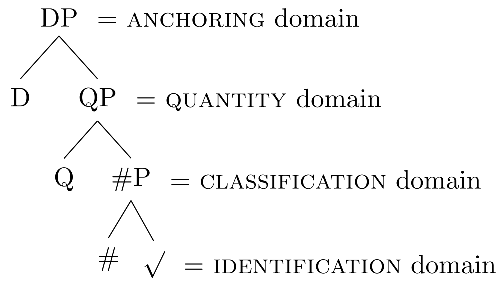
[9] I refer the reader to (Klockmann 2017:36–37) for detailed argumentation for this Universal Spine.

the classification domain. Klockmann (2017) proposes that this is the case for languages like Polish, which makes both number and gender distinctions. For Polish, the classification domain thus exists of both a number (#) and a gender (γ) projection (25).[10] In the English nominal phrase, the classification domain consists of only a number (#) projection, as this language makes no morphological distinction for gender (26).

(25) **Structure of Polish nominal phrases**



(26) **Structure of English nominal phrases**



Klockmann (2017:161) furthermore assumes that if languages can differ in the number of layers each domain has in the nominal phrase, they can also differ in the types of elements that can appear in these domains. For example, not all languages will necessarily use – or only use – a FVI bearing a [Q]-feature (like English *many*, cf. Table 2.1) in the quantity domain. Some also use other types of vocabulary items in this domain. It is also possible that multiple layers are used for the quantity domain. This brings us to her analysis of *bunch*, which involves lexicalising the quantity domain in English nominal phrases differently.[11]

[10] Klockmann (2017) assumes that gender is projected in syntax, right above the root. See Klockmann (2017:38) for a brief discussion of other approaches of where gender is encoded.

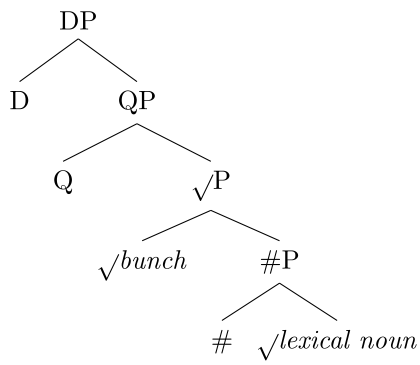
[11] For the purposes of this chapter, it suffices to illustrate Klockmann (2017)'s approach to semi-lexicality with her analysis of semi-lexical *bunch*. Note, however, that the way the syntactic features on semi-lexical roots influence their morphosyntactic environment is slightly

In the previous subsection we have seen that *bunch* has the following morphosyntactic properties when it is used semi-lexically. First, it always occurs with an indefinite article (*a bunch of chickens* versus **the bunch of chickens*). Second, it always occurs in the singular (*a bunch of chickens* versus **bunches of chickens*). Third, it indicates a quantity of the lexical noun. Based on her definition of semi-lexical items as roots specified for a (bundle of) syntactic feature(s), Klockmann (2017:268) proposes the lexical entry for semi-lexical *bunch* given in (27).[12]

(27) *Bunch*: [$\sqrt{bunch}$, Q, INDEF, $\neg\#_{pl}$]

The lexical entry of *bunch* consists of its root, a quantificational [Q]-feature, a [INDEF]-feature, and a negatively specified [$\#_{pl}$]-feature. Recall that Klockmann (2017) assumes that languages can differ in which elements are used to lexicalise a given domain in the nominal phrase. Given that *bunch* is specified for a [Q]-feature, it is a candidate to lexicalise the quantity domain of a nominal projection. This means that the semi-lexical root *bunch* is Merged above the classification domain (i.e. the number projection, cf. (26)) of the lexical root in a nominal phrase. However, because semi-lexical *bunch* is also still a root, it cannot be Merged directly in the Q-head, but it is Merged right below it. An illustration is given in (28).[13]

(28)



The [INDEF]-feature on the root of *bunch* furthermore ensures that it can only co-occur with an indefinite determiner, while the negatively specified [$\#_{pl}$]-feature blocks *bunch* from occurring in the plural.[14] Summing up, the [Q]-

different in the different case studies of semi-lexicality in her work.

[12] The specification Klockmann (2017:268) gives contains three additional features, which I have left out here as they do not have any impact on the syntactic structure of the nominal phrase in which *bunch* appears. These three features are [REL], [MAX], and [RANGE]. They encode that *bunch* is a relative quantifier, which indicates the upper range of a scale. See Klockmann (2017:250-251) for details and discussion.

[13] See Klockmann (2017:166-167) for further justification of this structure.

[14] I have simplified the analysis of the indefinite determiner and plural number here, as the

feature on the semi-lexical root of *bunch* makes it possible for this root to be Merged in the quantity domain of lexical nouns, right below the Q-head, and right above the classification domain. Two additional features account for the specific morphosyntactic behaviour of *bunch*, namely that it can only occur with an indefinite determiner, and never in the plural.[15]

We have thus seen that in Klockmann (2017)'s approach, semi-lexicality is not the result of a vocabulary item being used semi-lexically, but that semi-lexical vocabulary items exist independently and are marked as such in the lexicon. Semi-lexical vocabulary items are roots specified for one or more syntactic features. Klockmann furthermore assumes that nominal phrases consist of four domains, but that languages can differ in which elements lexicalise a given domain. Semi-lexical *bunch* is an example of how in English, the quantity domain can be lexicalised by a vocabulary item which is not a straightforward quantifier, but a root specified for a [Q]-feature. As she herself notes, this analysis raises the question of what the restrictions are on which types of elements can lexicalise a given domain in a given language. With respect to the quantity domain, she proposes that any element bearing a [Q]-feature, whether this is a root bearing the feature or just the feature itself, can occupy this domain (Klockmann 2017:161). In addition, she also briefly explores how a lexical vocabulary item could acquire such a [Q]-feature diachronically, and thus grammaticalise from a lexical vocabulary item (a root) to a semi-lexical vocabulary item (a root with a [Q]-feature). The first step in the grammaticalisation of a lexical vocabulary item into a semi-lexical one, involves using this lexical vocabulary in a "more functional" context, for example in one of the domains above the identification domain (i.e. the position of the lexical root). As she phrases it:

> The idea [is] that by using a particular root in a more functional context, it can acquire a functional flavor without necessarily being specified for functional features. This is presumably the first step in processes of grammaticalisation (Klockmann 2017:268).

Certain conditions have to be met for a lexical vocabulary item to be used in the functional domain of another root, however. First, the lexical semantics should be (at least partly) compatible with the functional domain in which it is going to be used. For example, consider the semi-lexical uses of a number of lexical vocabulary items expressing quantity:

(29) a. A **wealth** of examples
b. A **flood** of memories
c. A **parade** of witnesses (Klockmann 2017:268)

specific technical details would take me too far afield. The interested reader is referred to (Klockmann 2017:26–264).

[15] Note that in the structure in (28), there is no syntactic position for *of*, which obligatorily occurs between *bunch* and the lexical noun. Klockmann (2017:271-272) hypothesises that *of* is an element which indicates the presence of a classification domain.

Wealth, *flood* and *parade* all have lexical semantics involving the notion of a quantity (specifically in these cases a large quantity). Nouns of which the lexical semantics do not contain a reference to a type of quantity cannot be used in the quantity domain of another noun. For example, the noun *honesty* cannot indicate the quantity of a lexical noun, given that its lexical semantics does not refer to a quantity in any way:

(30) *An **honesty** of examples/memories/witnesses (Klockmann 2017:268)

Second, the language should already make use of a specific morphosyntactic structure, in which a lexical vocabulary item with the right lexical semantics can be inserted (Klockmann 2017:269). For example, because English already has a morphosyntactic construction in which quantificational nouns like *bunch* (but also *lot* and *number*) take an indefinite determiner and are separated from the lexical noun by *of*, lexical vocabulary items such as *wealth*, *flood*, and *parade* can be inserted into the same morphosyntactic construction.

Summing up, the first step of the grammaticalisation path from lexical vocabulary item (a root) to semi-lexical vocabulary item (a root specified for a syntactic feature) involves the use of that lexical vocabulary item in the functional domain of another root (the lexical noun of the nominal phrase). In order for this to happen, two conditions have to be met. First, the lexical vocabulary item that will be used semi-lexically has lexical semantics that are compatible with the domain in which it will be inserted. Second, there already is a morphosyntactic template available in the language that roots can be used in.

An important question regarding this grammaticalisation process is how the lexical vocabulary item that is used semi-lexically (like *wealth*) can eventually acquire a syntactic feature (a [Q]-feature), which will make it an actual semi-lexical vocabulary item. Klockmann (2017) does not make explicit the step from a lexical vocabulary item being used in the quantity domain of a root of a nominal phrase to this vocabulary item being specified for a [Q]-feature in the lexicon.[16]

In my approach to semi-lexicality I make an explicit proposal about what the early stages of grammaticalisation look like, and specifically about the underlying syntactic structures of these first stages in the verbal domain of Dutch and Afrikaans. I build on Klockmann (2017)'s proposal that semi-lexicality involves a root that is being used in the functional domain of another root (see the structure in (28)). However, what I do not adopt from her approach is the

[16]Note also that semi-lexical *bunch* is not only specified for a [Q]-feature, but also for a [INDEF]- and a $\neg[\#_{pl}]$-feature, as we have seen above (cf. the lexical entry for *bunch* in (27)). Klockmann (2017) does not discuss how *bunch* acquired these features in the course of its grammaticalisation process. She is more explicit about the grammaticalisation path of Polish semi-lexical nouns that are used as numerals, see (Klockmann 2017:chapter 4–5). However, even there she is only explicit about the later stages of grammaticalisation, which involve feature loss (the loss of gender and number features), rather than the first steps, in which these semi-lexical nouns have acquired their [Q]-feature.

idea that there exist actual semi-lexical items in the lexicon. That is, I keep to the dichotomy between lexical versus functional vocabulary items, in which the former spell out roots, and the latter a (bundle of) syntactic feature(s). Before I present my proposal, I first need to provide more background about semi-lexicality in the verbal domain in Dutch and Afrikaans, namely semi-lexical restructuring. This is the topic of the next section.

2.3 Semi-lexicality in the verbal domain: restructuring

In this section, I discuss the contexts in which semi-lexicality in the Dutch and Afrikaans verbal domain arises, namely those involving restructuring. In subsection 2.3.1, I introduce the phenomenon of restructuring. In subsection 2.3.2, I discuss Wurmbrand (2001)'s proposal to distinguish cases of lexical restructuring from cases of functional restructuring. In subsection 2.3.3, I turn to semi-lexical restructuring, and discuss Wurmbrand (2001)'s analysis of this. Furthermore, I present Dutch data that pose a problem for her analysis of semi-lexical structuring.

2.3.1 Restructuring: the phenomenon

Restructuring is a phenomenon which can be observed when certain types of verbs (modal, aspectual, motion and causative verbs, but also other verbs, like *try*, *dare* et cetera) select an infinitive, and the result is a mono-clausal structure.[17] The fact that the structure is mono-clausal is evidenced by so-called restructuring phenomena or transparency properties. For Romance languages, those phenomena include clitic climbing, auxiliary switch and object preposing. In Germanic languages, they include long distance scrambling, long passive and verb raising (Wurmbrand 2001).[18,19] An example of clitic climbing in Italian is given in (31-a).

(31) a. ***Ti*** **vorrei** baciare.
You would.want kiss.
'I would like to kiss you.'

b. **Vorrei** baciar-***ti***.
would.want kiss-you.
'I would like to kiss you.'

[17]See Wurmbrand (2001:10–13) for a discussion of different types of analyses of this mono-clausal structure.

[18]Auxiliary switch is attested in Dutch restructuring contexts as well (Haeseryn et al. 1997, Van Eynde et al. 2016). The reader is referred to Ter Beek (2008) for a detailed discussion of restructuring phenomena in Dutch.

[19]See Wurmbrand (2015) for a typological overview of restructuring.

(32) a. * ***Ti*** **detesterei** baciare.
You would.hate kiss

b. **Detesterei** baciar-***ti***.
would.hate kiss-you
'I would hate to kiss you.'

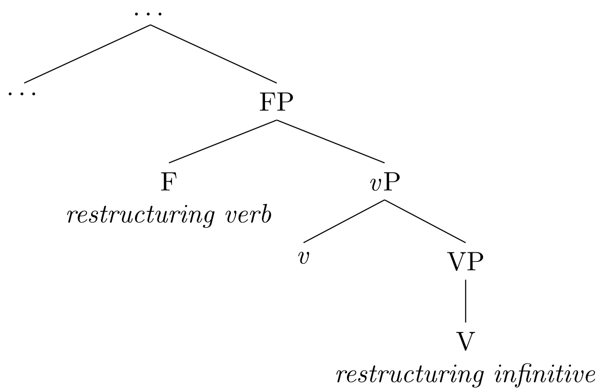
In (31-a), the matrix verb is a modal verb (*vorrei* 'would want'), and the object clitic (*ti* 'you'), which is the object of the embedded infinitive (*baciare* 'kiss'), appears on the modal verb rather than the infinitive. In other words, the clitic *ti* has climbed up from its lower position on the infinitive to a higher position on the modal verb. Clitic climbing is optional, as can be seen in (31-b). When the matrix verb is a lexical rather than a functional verb, clitic climbing is blocked. This is shown in (32-a) with the matrix verb *detesterei* 'would hate'. The clitic has to remain on the embedded infinitive (32-b).

In the next subsections, all illustrations of restructuring come from Germanic, as this thesis is concerned with semi-lexical restructuring in two Germanic languages (Dutch and Afrikaans). As my analysis of semi-lexical restructuring in these languages builds on the analysis of Wurmbrand (2001), her work is discussed in detail as well, mainly as it pertains to cases of German restructuring. The following labels are used throughout the rest of the section. The verb that selects an infinitive (e.g. the modal *vorrei* 'would want' in (31)) is referred to as the *restructuring verb*. The selected infinitive (e.g. *baciare* 'kiss' in (31)) is referred to as the *restructuring infinitive*. The combination of the restructuring verb and the restructuring infinitive is called the *restructuring configuration*.

2.3.2 Lexical and functional restructuring

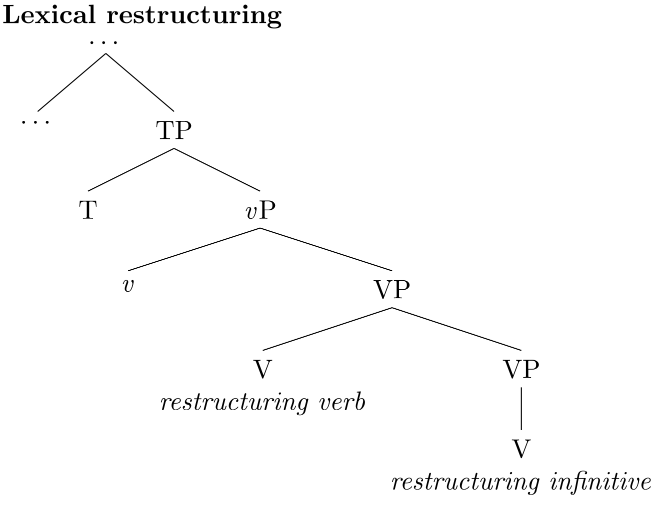
An influential approach to restructuring is the one by Cinque (2001, 2003, 2006). He argues that in all restructuring configurations, the restructuring verbs (modals, aspectual verbs, motion verbs, and verbs like *try* or *dare*) are functional heads in the functional domain of the restructuring infinitive. The gist of Cinque's analysis can be represented as in (33) (based on Wurmbrand (2004:992)).

(33) **Functional restructuring**



As can be seen in the structure, the restructuring verb is in a functional head F in the functional domain of the restructuring infinitive in V. Wurmbrand (2004)) argues against the idea that all cases of restructuring are of the type in (33), and proposes instead that there are two types of restructuring: functional and lexical restructuring. The structure in (33) represents a functional restructuring configuration. The core class of functional restructuring verbs consists of modals and auxiliaries. In cases of lexical restructuring, the restructuring verb is fully lexical, and thus cannot be merged in a functional head. The restructuring infinitive is lexical as well, but its 'clause' is very limited in size (e.g. a VP). A lexical restructuring configuration is illustrated in (34).

(34) **Lexical restructuring**



As can be seen in the structure, the restructuring verb heads its own VP in lexical restructuring configurations. The restructuring infinitive does so too, and it is a complement of the lexical restructuring verb, albeit one that is very small in size (only a VP). The class of lexical restructuring verbs includes verbs like *try*, *begin*, and *dare*.[20]

Wurmbrand (2001, 2004) bases her argumentation for a split between functional and lexical restructuring verbs on the fact that in German, these verbs systematically show distinct morphosyntactic behaviour. Three morphosyntactic differences between the two classes of restructuring verbs are (i) their thematic restrictions, (ii) the morphological form in which they occur when embedded under a perfective auxiliary, and (iii) whether or not the restructuring verb allows extraposition of the embedded infinitive. I illustrate all three differences in turn.

One clear difference in thematic restrictions between functional and lexical restructuring verbs in German is that the former allow weather-*it* subjects, whereas the latter do not.[21] This is illustrated in (35) for a number of functional restructuring verbs, and in (36) for a number of lexical restructuring verbs.

(35) *Functional restructuring verbs*

a. ***Es*** **muß** schneien.
   it must snow
   'It has to snow.'

b. ***Es*** **dürfte** schneien.
   it might snow
   'It might snow.' (Wurmbrand 2001:169)

(36) *Lexical restructuring verbs*

a. ****Es*** **versucht** zu schneien.
   it tries to snow

b. ****Es*** **wagt** zu schneien.
   it dares to snow

(Wurmbrand 2001:168)

As can be seen in (35), functional restructuring verbs can readily be combined with a weather-*it* subject, whereas this is blocked with lexical restructuring verbs (36).

A second difference in the morphosyntactic behaviour of functional and lexical restructuring verbs concerns the so-called IPP-effect (IPP stands for *Infinitivus Pro Participio*), a peculiar property found in certain restructuring contexts in West Germanic languages. Normally, when a verb is embedded under a perfective auxiliary, it surfaces as a participle. In the Dutch sentence in (37), for example, the perfective auxiliary *heb* 'have' selects the modal perfect participle *gewild* 'wanted'.

[20] For the full list of lexical and functional restructuring verbs in German, see Wurmbrand (2001).

[21] Note that there are other ways to see the difference between functional and lexical restructuring verbs with respect to their thematic restrictions. The interested reader is referred to Wurmbrand (2001:chapter 2–3) and Wurmbrand (2004:993–996) for extensive illustrations and discussion.

(37) Ik **heb** dit zo ***gewild***.
I have this so wanted.PTCP
'I have wanted it this way.'

However, when the modal itself selects an infinitive, as in (38), it surfaces in its infinitival form *willen* 'want', not as a participle.[22]

(38) Ik heb dit zo *__gewild__ / **willen** doen.
I have this so want.PTCP / want.IPP do
'I wanted to do it this way.'

Wurmbrand (2001) shows that in German there is a clear divide between functional and lexical restructuring verbs when it comes to the IPP-effect. Functional restructuring verbs show the effect, as illustrated for two modal verbs in (39). The perfect participle form is allowed in certain dialects, though, as indicated by the percentage sign before the perfect participle. In contrast, lexical restructuring verbs never show the IPP-effect, as illustrated in (40).[23]

(39) a. Hans hat nach Hause gehen **muß** / %**gemußt**.
Hans has to home go must.IPP / must.PTCP
'Hans had to go home.'
b. Hans hat nach Hause gehen **dürfen** / %**gedurft**.
Hans has to home go might.IPP / might.PTCP
'Hans was allowed to go home.' (Wurmbrand 2001:165)

(40) a. Hans hat nach Hause zu gehen *__versuchen__ / **versucht**.
Hans has to home to go try.IPP / try.PTCP
'Hans has tried to go home.'
b. Hans hat nach Hause zu gehen *__wagen__ / **gewagt**.
Hans has to home to go dare.IPP / dare.PTCP
'Hans has dare to go home.' (Wurmbrand 2001:163)

The third difference between functional and lexical restructuring verbs is that extraposition of the embedded infinitive is only possible with the latter, not the former. That is, the infinitival complement (indicated by means of square brackets in the examples below) cannot appear linearly to the right of functional

[22] It is not clear whether the form of the modal is a real infinitival form here, or whether it only resembles such a form (for discussion, see Hinterhölzl 2009, Schmid 2005, Zwart 2007). For example, if the verb undergoing IPP is the auxiliary *zijn* 'to be', the form it surfaces in is distinct from its regular infinitival form:

(i) Ik **ben** *___zijn___ / ***wezen*** vissen.
I am be.INF / be.IPP fish
'I went fishing.' (Zwart 2007:78)

Furthermore, in many Dutch dialects, the IPP form is distinct from the regular infinitival form for other restructuring verbs as well, see Barbiers et al. (2008).

[23] As Wurmbrand herself notes, the division is not as clear-cut in Dutch, given that a number of the verbs she considers to be lexical restructuring verbs nonetheless optionally show the IPP-effect, for example the Dutch verb *proberen* 'try' (Haeseryn et al. 1997).

restructuring verbs (41), whereas it can do so with lexical restructuring verbs (42).[24]

(41) a. *. . . dass Hans **muß** [den Kuchen essen].
. . . that Hans must the cake eat
b. *. . . dass Hans **dürfte** [den Kuchen essen].
. . . that Hans might the cake eat
(Wurmbrand 2001:157)

(42) a. . . . dass Hans **versucht** [den Kuchen zu essen].
. . . that Hans tries the cake to eat
'. . . that Hans tries to eat the cake.'
b. . . . das Hans **wagt** [den Kuchen zu essen].
. . . that Hans dares the cake to eat
'. . . that Hans dares to eat the cake.' (Wurmbrand 2001:156)

The moprhosyntactic differences between functional and lexical restructuring verbs are summarised in Table 2.2.

Restructuring verb	Thematic properties	IPP-effect	Extraposition
Functional	NO	YES	NO
Lexical	YES	NO	YES

Table 2.2: Properties of functional and lexical restructuring verbs

2.3.3 Semi-lexical restructuring

A problem for the functional versus lexical restructuring divide is the fact that there are a number of verbs the syntactic behaviour of which is 'in between' that of functional and lexical restructuring verbs. Wurmbrand (2001) therefore argues, following Riemsdijk (1998b), that the functional/lexical distinction does not suffice, and that there is also a class of semi-lexical restructuring verbs. In German, this class consists of the following verbs: the motion verbs *gehen* 'go' and *kommen* 'come', the perception verbs *hören* 'hear' and *sehen* 'see', and causative *lassen* 'let'.[25] With respect to their thematic properties, these verbs behave like lexical restructuring verbs, in that they do not allow weather-*it* subjects. This is illustrated in (43).

[24]The (im)possibility of extraposition of the embedded infinitive is not influenced by the presence or absence of the infinitival marker *zu*, see Wurmbrand (2001:161).

[25]Wurmbrand also places two of the modals, namely *können* 'can' and *wollen* 'want', in the class of semi-lexical restructuring verbs, but adds that they are ambiguous between semi-lexical and functional restructuring verbs. Given that she does not focus on these two modals in her discussion of semi-lexical verbs, I leave them aside here as well.

(43) *Semi-lexical restructuring verbs*

a. ****Es*** **geht** schneien.
it goes snow

b. *Es **kommt** schneien.
it comes snow

c. ****Es*** **hört** den Peter donnern.
it hears the Peter thunder

d. ****Es*** **sah** den Peter schneien.
it saw the Peter snow

e. ****Es*** **lässt** den Peter donnern.
it let the Peter thunder (Wurmbrand 2001:171)

Regarding the IPP-effect, three of the semi-lexical restructuring verbs show optional IPP (44).

(44) a. Hans hat den Peter musizieren **gehört** / **hören**.
Hans has the Peter make.music hear.PTCP / hear.IPP
'Hans has heard Peter make music.'

b. Hans hat den Peter musizieren **gelassen** / **lassen**.
Hans has the Peter make.music let.PTCP / let.IPP
'Hans has let Peter make music.'

c. Hans hat den Peter musizieren **gesehen** / **sehen**.
Hans has the Peter make.music see.PTCP / see.IPP
'Hans has seen Peter make music.' (Wurmbrand 2001:165)

For the other semi-lexical verbs, *gehen* 'go', and *kommen* 'come', the situation is more complicated. In a nutshell, these verbs do not show the IPP-effect, but Wurmbrand (2001) argues that this is due to the fact that the perfective auxiliary that selects them is *sein* 'be' rather than *haben* 'have'. The reader is referred to Wurmbrand (2001:162–164) for full discussion and illustration. We can thus conclude that for these two verbs, the IPP-test is not applicable.

Finally, with respect to extraposition, the semi-lexical restructuring verbs behave like functional restructuring verbs in not allowing it.

(45) a. *. . . dass Hans **geht** [den Kuchen essen].
. . . that Hans goes the case eat

b. *. . . dass Hans **hörte** [den Peter den Kuchen essen].
. . . that Hans heard the Peter the cake eat

c. *. . . dass Hans **kommt** [den Kuchen essen].
. . . that Hans comes the cake eat

d. *. . . dass Hans **ließ** [den Peter den Kuchen essen].
. . . that Hans let the Peter the cake eat

e. *. . . dass Hans **sah** [den Peter den Kuchen essen].
. . . that Hans saw the Peter the cake eat

(Wurmbrand 2001:158)

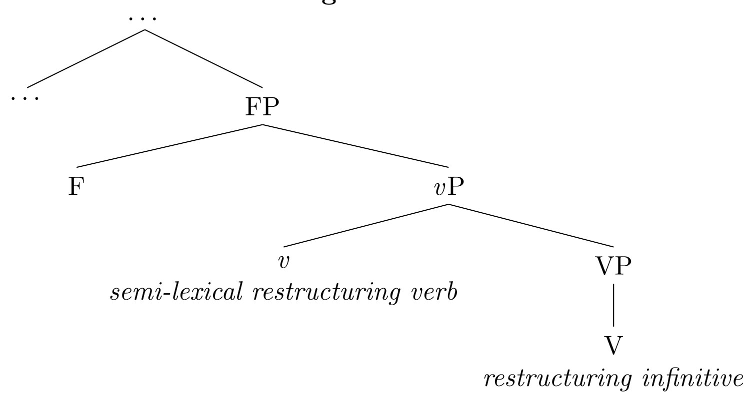
Summing up, we have seen that in German there is reason to assume the existence of a third class of restructuring verbs: semi-lexical ones. These verbs pattern together with functional restructuring verbs in not allowing extraposition, they behave 'in between' the two classes of verbs with respect to the IPP-effect, and they pattern together with lexical restructuring verbs in displaying thematic restrictions. These findings are summarised in Table 2.3.

Restructuring verb	Thematic properties	IPP-effect	Extraposition
Functional	NO	YES	NO
Semi-lexical	YES	N.A./OPTIONAL	NO
Lexical	YES	NO	YES

Table 2.3: Properties of functional, semi-lexical and lexical restructuring verbs

Wurmbrand (2001)'s analysis of semi-lexical verbs is that they are voice/aspect elements, and that they are inserted in v/Asp above the lexical verb (she does not make a distinction between v and Asp for German and calls this head v). This is illustrated in the structure in (46).

(46) **Semi-lexical restructuring**



Wurmbrand (2001) proposal that all semi-lexical verbs are merged in v in German is supported by three pieces of data. First, these verbs can be embedded under modals but they cannot themselves embed a modal. Second, they cannot embed nor be embedded by other semi-lexical verbs. Third, they cannot be passivised, nor can they embed a passive predicate. These facts are illustrated for the perception verb *sehen* 'see' in the examples below, but the same holds for causatives and motion verbs as well.

(47) a. Hans **darf** Maria nicht musizieren **sehen**.
Hans may Maria not make.music see
'Hans may not see Mary make music.'

b. *Hans **sah** den Peter musizieren **dürfen**.
Hans saw the Peter make.music may
(Wurmbrand 2001:218)

In (47-a), we see that a modal (*darf* 'may') can embed the semi-lexical verb **sehen** 'see', whereas the reverse is not possible (47-b). Modals are functional restructuring verbs and thus assumed to be situated in F. The fact that they can embed semi-lexical restructuring verbs, but that semi-lexical restructuring verbs cannot embed modals follows from the structure in (46), in which the restructuring verb is situated in a structurally lower position than the modal.

(48) a. *Hans **geht** Peter musizieren **sehen**.
Hans goes Peter make.music see
b. *Hans **sah** ihn einen Turm bauen **gehen**.
Hans saw him a tower build go
(Wurmbrand 2001:219)

In (48-a), we see that the semi-lexical restructuring verb *gehen* 'go' cannot embed the semi-lexical restructuring verb *sehen* 'see'. The reverse is also ungrammatical, as can be seen in (48-b). These facts support the analysis in which all semi-lexical restructuring verbs occur in the same syntactic position.

(49) a. *Der Peter **wurde** den Kuchen essen **gesehen**.
The Peter was.PASS the cake eat seen
b. *Hans **sah** den Kuchen gegessen **werden**.
Hans saw the cake eat been.PASS
(Wurmbrand 2001:220-221)

Finally, in (49-a), we see that the semi-lexical restructuring verb *sehen* 'see' cannot be passivised. In (49-b), we furthermore see that *sehen* 'see' cannot embed a passive. If we assume that v is connected to Voice, these facts support Wurmbrand (2001)'s analysis of semi-lexical verbs as being Merged in v. That is, if v is already occupied by a semi-lexical verb, a passive construction is excluded, and vice versa.

A problem for Wurmbrand (2001)'s analysis is the fact that in Dutch, there is a group of aspectual verbs that are semi-lexical restructuring verbs based on her tests, but that can be embedded by other Dutch semi-lexical restructuring verbs. These verbs are the motion verb *lopen* 'walk', and the three posture verbs *zitten* 'sit', *staan* 'stand' and *liggen* 'lie'. In (50), I show that all these verbs reject weather-*it* subjects, and thus behave like lexical restructuring verbs. In (51), I show that they reject extraposition, and thus behave like functional restructuring verbs.[26]

[26] Unlike the German semi-lexical restructuring verbs, these verbs show obligatory IPP when embedded under a temporal auxiliary. Dutch has much more IPP verbs than German in general, though, see Schmid (2005).

(50) a. ****Het*** **loopt** te sneeuwen.
it walks to snow
b. ****Het*** **zit** te sneeuwen.
it sits to snow
c. ****Het*** **staat** te sneeuwen.
it stands to snow
d. ****Het*** **ligt** te sneeuwen.
it lies to snow

(51) a. *…dat ik **loop** [een koekje *te eten*].
…that I walk a cookie to eat
b. *…dat ik **zit** [een koekje *te eten*].
…that I sit a cookie to eat
c. *…dat ik **sta** [een koekje *te eten*].
…that I stand a cookie to eat
d. *…dat ik **lig** [een koekje *te eten*].
…that I lie a cookie to eat

These verbs can be embedded both by modal verbs and by other semi-lexical verbs (see also Van Craenenbroeck and Van Koppen (2017)), but they cannot embed modals or semi-lexical verbs themselves. This is illustrated for the modal *moeten* 'must', the semi-lexical verb *zien* 'see', and the semi-lexical verb *zitten* 'sit' in (52) and (53), but it holds for the motion verb *lopen* and the other two posture verbs as well.

(52) a. Ik **moet** de hele dag **zitten** te werken.
I must the entire day sit to work
'I have to be working the entire day.'
b. *Ik **zit** de hele dag te **moeten** werken.
I sit the entire day to must work

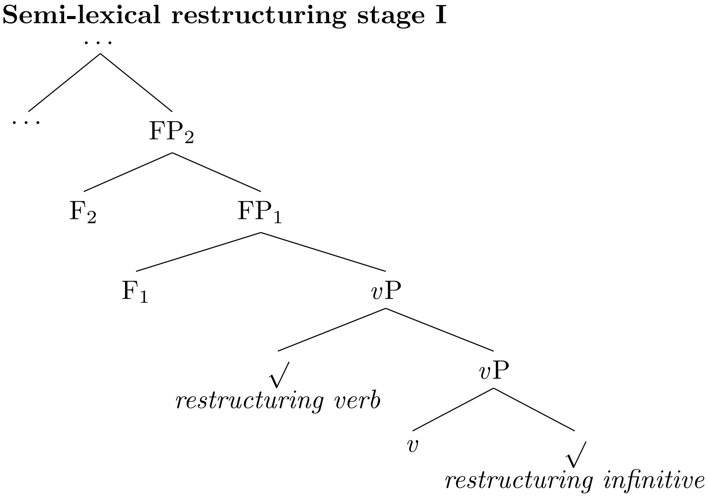
(53) a. Ik **zie** hem de hele dag **zitten** te werken.
I see him the entire day sit to work
'I see him working the entire day.'
b. *Ik *zit* hem de hele dag te **zien** werken.
I sit him the entire day to see work

The fact that these verbs can be embedded under both modals and other semi-lexical verbs, means that they must be situated in a position which is structurally lower than both the position of modals and that of the other semi-lexical verbs. This means that Wurmbrand (2001)'s analysis of German, in which all semi-lexical restructuring verbs are situated in v, cannot be extended to Dutch. For the Dutch aspectual semi-lexical restructuring verbs, a different analysis is needed. This leads me to the next section, in which I present my proposal for semi-lexical restructuring.

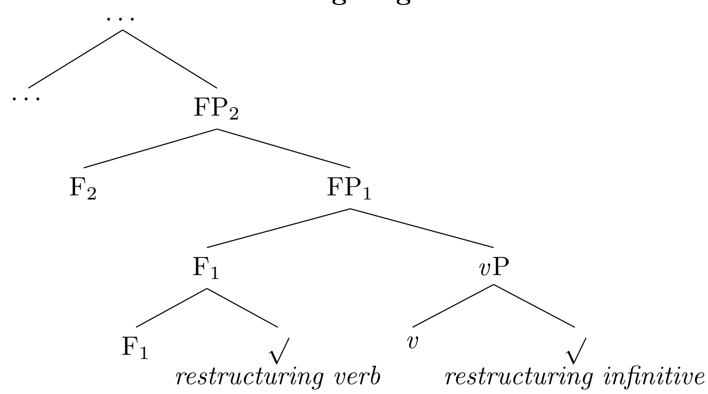
2.4 The proposal

In this section, I present my proposal of semi-lexical restructuring. As already mentioned at the end of subsection 2.2.2, my aim is to make a concrete proposal for the very early steps of grammaticalisation, i.e. the grammaticalisation path from lexical vocabulary item to lexical vocabulary item that can also be used semi-lexically. I propose that this very early part of grammaticalisation proceeds in two stages, the underlying syntactic structures of which are distinct from each other. Specifically for restructuring, the proposed structure for the first stage of semi-lexicality is given in (54), and for the second stage in (55).

(54) **Semi-lexical restructuring stage I**



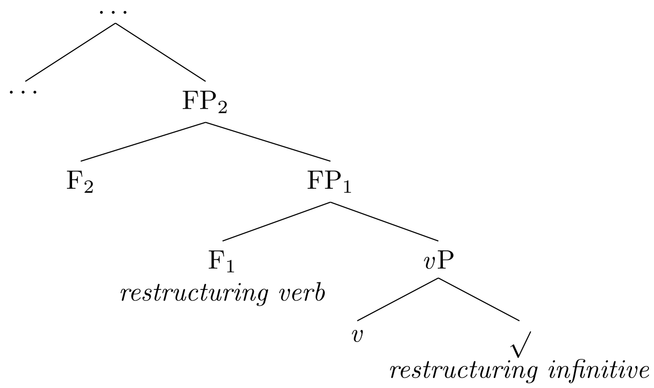
(55) **Semi-lexical restructuring stage II**



In (54)– the first stage – the restructuring verb verb is a root which is Merged in the (very) low regions of the functional domain of the restructuring infinitive (cf. Klockmann (2017)'s structure for the quantificational noun *bunch* in subsection 2.2.2). In (55)– the second stage – the restructuring verb is still a root, but it is Merged with a functional head (i.e. a syntactic feature) in a separate workspace. This complex head is then Merged into the functional domain of the restructuring infinitive (see also Song (2019), who analyses semi-lexicality as involving a root that forms a complex head with a functional head, and Biberauer (2016:3)). When and how a semi-lexical restructuring verb can grammaticalise from the first to the second stage is made more explicit in the two case studies that follow this chapter.

For functional restructuring I adopt the analysis proposed by Wurmbrand (2001), in which the restructuring verb is Merged in a functional head in the functional domain of the restructuring infinitive (56).

(56) **Functional restructuring**



Important to note here is the fact that the restructuring verb is no longer a root, but a syntactic feature. The functional restructuring stage is the next step in the grammaticalisation process after the second stage of semi-lexicality (i.e. (55)). Note, however, that the grammaticalisation can halt at any of these stages: once a grammaticalisation process is started, there is nothing that forces it to be fully completed (see among others Hopper and Traugott (1993)).

With the structures of the two stages of semi-lexical restructuring in place, I move on to the two case studies, in which semi-lexical restructuring in Dutch and Afrikaans is investigated in depth, and in which the proposal is fleshed out in more detail.

Case study I

Semi-lexicality in Dutch non-finite verb clusters

CHAPTER 3

Introduction

Onder mijn boot zie
ik een donkere schaduw
en huiver verrukt

- Peter Pots

In recent years, a lot of insight has been gained into the variation – as well as the limits on that variation – in the word order of Germanic finite verb clusters, i.e. verb clusters in which the finite verb of the sentence is part of the cluster (Barbiers et al. 2008, Wurmbrand 2017). An example of the attested word order variation in a Dutch verb cluster with three verbs in the configuration *modal–modal–infinitive* is given in (1). As is standard in the verb cluster literature, I refer to the hierarchical order of the verbs with subscript numbers: V_1 selects V_2 and V_2 selects V_3. In finite verb clusters, V_1 is always the finite verb.

(1)	a.	Ik	vind	dat	iedereen	**moet**$_1$	**kunnen**$_2$	**zwemmen**$_3$.	(1-2-3)
		I	find	that	everyone	must.FIN	can.INF	swim.INF	
	b.	Ik	vind	dat	iedereen	**moet**$_1$	**zwemmen**$_3$	**kunnen**$_2$.	(1-3-2)
		I	find	that	everyone	must.FIN	swim.INF	can.INF	
	c.	Ik	vind	dat	iedereen	**zwemmen**$_3$	**kunnen**$_2$	**moet**$_1$.	(3-2-1)
		I	find	that	everyone	swim.INF	can.INF	must.FIN	
	d.	Ik	vind	dat	iedereen	**zwemmen**$_3$	**moet**$_1$	**kunnen**$_2$.	(3-1-2)
		I	find	that	everyone	swim.INF	must.FIN	can.INF	
	e.	*Ik	vind	dat	iedereen	**kunnen**$_2$	**moet**$_1$	**zwemmen**$_3$.	(2-1-3)
		I	find	that	everyone	can.INF	must.FIN	swim.INF	

f. *Ik vind dat iedereen **kunnen**$_2$ **zwemmen**$_3$ **moet**$_1$. (2-3-1)
I find that everyone can.INF swim.INF must.FIN
'I feel everyone should be able to swim.' (Barbiers and Bennis 2010:26)

As can be seen in this example, in a three-verb cluster with two modals and an infinitive, four out of the six logically possible word orders are grammatical, and two are not (2-1-3 and 2-3-1). It is not the case that all Dutch varieties allow all four word orders of the cluster, though: there is substantial geographical variation in these clusters. Another factor that influences the variation is the types of verbs making up the cluster. For example, the 2-3-1 order, which is ungrammatical with a verb cluster with two modals (e.g. (1-f)), is grammatical in a cluster with an auxiliary, an aspectual or modal verb and an infinitive (*aux-asp/modal-infinitive*). In this type of cluster, the 1-3-2 order does not occur, however, while it is very frequent in the *modal-modal-infinitive* cluster illustrated in (1). Still other frequency and geographical patterns occur in a cluster in which a modal selects a perfective auxiliary (*modal-aux-participle*), et cetera. In other words, as Barbiers and Bennis (2010) show, at least three factors interactively affect the word order variation (and its limits) in Dutch finite three-verb clusters, namely (i) geography, (ii) the category of the verbs making the cluster, and (iii) the hierarchical relations of these verbs. For the most recent syntactic analysis in a generative framework of Dutch verb cluster variation, see Barbiers et al. (2018) and Dros-Hendriks (2018), and for a statistically based evaluation of their analysis, see Van Craenenbroeck (2019).

In contrast to the extensive body of work on word order variation in finite verb clusters in Dutch (and more broadly Germanic verb clustering varieties: Dutch, German, Swiss, Frisian and Afrikaans and their dialects), hardly any attention has been paid to variation in non-finite verb clusters, neither empirically nor theoretically. Non-finite verb clusters are clusters that consist of only non-finite verbs, i.e. infinitives or participles. In the case of non-finite clusters, the finite verb – if there is one – is positioned in Verb Second position, and is thus not part of the cluster. That is, it does not participate in word order variation. In this case study, I focus on Dutch non-finite verb clusters that consist of three infinitives. An example of such a cluster is given in bold in (2).

(2) Koen *zal* morgen niet **hoeven**$_1$ **te gaan**$_2$ **werken**$_3$.
Koen will.FIN tomorrow not need.INF to go.INF werken.INF
'Koen won't have to go and work tomorrow.'

This chapter focusses on variation in the placement and presence of *te* 'to', a morpheme that appears in verb clusters as a result of the selectional requirements of one of the verbs.[1,2] For example, in (2), V_1 *hoeven* 'need' selects a *te*-infinitive, which means that *te* surfaces on V_2 *gaan* 'go'.

In this case study I show that there is an extensive amount of variation regarding the presence and placement of *te*. For example, even though the selectional requirements dictate that *te* surface on V_2 in (2), many Dutch speakers allow it to occur on V_1 (3-a). A smaller group of speakers allow *te* to occur on V_3 (3-b), or for it to occur twice in the same cluster (3-c).[3] A large group of speakers also allows for *te* to be absent altogether (3-d). In all examples, the verb on which *te* should appear based on selection requirements is given in bold.

(3) a. Koen *zal* morgen niet [***te* hoeven**$_1$ gaan$_2$ werken$_3$].
Koen will.FIN tomorrow not to need.INF go.INF work.INF
b. Koen *zal* morgen niet [hoeven$_1$ **gaan**$_2$ ***te*** werken$_3$].
Koen will.FIN tomorrow not need.INF go.INF to work.INF
c. Koen *zal* morgen niet [***te*** hoeven$_1$ ***te* gaan**$_2$ werken$_3$].
Koen will.FIN tomorrow not to need.INF to go.INF work.INF
d. Koen *zal* morgen niet [hoeven$_1$ **gaan**$_2$ werken$_3$].
Koen will.FIN tomorrow not need.INF go.INF work.INF

Even though it has been noted in Meussen (1943), Vanacker (1969), Zwart (1993) and Dreumel and Coppen (2003) that *te* can appear on a hierarchically lower verb than it should based on selectional requirements, it has never been systematically investigated throughout the entire Dutch language area. The same holds for its absence in the cluster when selected by the two verbs that form the focus of this case study, namely *hoeven* 'need' and *zitten* 'sit' (De Rooij 1981, Haeseryn et al. 1997, Zwart 1993, Van Pottelberge 2002, Dreumel and Coppen 2003, Van de Velde 2017). That *te* can also appear on a hierarchically higher verb than the one it should appear on, and that it can be doubled, are

[1] I gloss *te* here and in the examples throughout this chapter as English 'to'. This is for ease of exposition, however, and does not mean they should receive the same morphosyntactic analysis. As will be clear from the data presented in this case study, it is even very unlikely that they should be analysed in the same way, as Dutch *te* shows much more variation in its presence and position than English 'to'. See also Wurmbrand (2001) on why English *to* and German *zu* cannot be analysed in the same way.

[2] I only investigate variation in the presence and placement of *te* when it is selected by one of the verbs in a clause, i.e. sentences like (2). Note, though, that *te* can also be selected by the non-finite complementiser *om* 'for', as illustrated in (i).

(i) Het is vervelend ***om*** lang **te** moeten wachten.
It is annoying for long to must wait
'It's annoying having to wait very long.'

Investigating variation in the presence and placement of *te* when selected by *om* is beyond the scope of this case study, however, and is left for future research.

[3] Given that there are three verbs in the cluster, there are three theoretically possible doubling configurations: *te*-V_1-*te*-V_2-V_3, *te*-V_1-V_2-*te*-V_3 and V_1-*te*-V_2-*te*-V_3. As will be shown in section 4.3.4, all three configurations are attested, albeit with different frequencies.

new empirical facts.[4] Moreover, the displacement patterns of *te* are systematically different from *zu*-displacement in standard German, and Alemannic and Swiss varieties, as discussed in Salzmann (2013, 2016, 2019a). In German, *zu*-displacement happens in verb clusters with a (partly) ascending verb cluster order, for example 1-3-2 or 3-1-2, rather than in the standard German word order, which is strictly descending (3-2-1) (though see Schallert (2019) for some exceptions to Salzmann's generalizations). Consider four examples from Standard German in (4-a)-(4-d).

(4) a. …ohne das Buch lesen$_3$ gekonnt$_2$ ***zu* haben**$_1$. (3-2-1)
…without the book read.INF can.PTCP to have.INF

b. …ohne das Buch haben$_1$ lesen$_3$ ***zu* können**$_2$. (1-3-2)
…without the book have.IPP read.INF to can.PTCP

c. …ohne das Buch lesen$_3$ haben$_1$ ***zu* können**$_2$. (3-1-2)
…without the book read.INF have.IPP to can.PTCP

d. *…ohne das Buch ***zu* haben**$_1$ lesen$_3$ können$_2$. (1-3-2)
…without the book to have.IPP read.INF can.PTCP
'…without having been able to read the book.' (Salzmann 2016:405–406)

In all these examples, *ohne* 'without' selects a *zu*-infinitive, which means that *zu* should appear on the hierarchically highest verb in the cluster, V_1 *haben* 'have'. This is indeed what we see in (4-a). However, when the order of the verb cluster deviates from the standard 3-2-1 order, *zu* no longer appears on the hierarchically highest verb of the cluster (V_1), but on V_2 *können* 'can' (4-b)–(4-c). Importantly, in a non-strictly-descending order (e.g. 1-3-2), the 'correct' placement of *zu* on V_1 is ungrammatical (4-d). *Zu*-displacement in non-strictly-descending word orders is thus obligatory. As we will see in section 4.3, this differs sharply from *te*-displacement in Dutch, which is almost always optional. A second difference between *zu*-displacement and *te*-displacement is that the former seems to be sensitive to linear order rather than hierarchical structure: when *zu* is selected by an element like *ohne* 'without' outside of the cluster, it has to appear on the linearly right-most verb of the cluster (i.e. on V_1 in the 3-2-1 order, and on V_2 in the 1-3-2 or 3-1-2 order). In this chapter, we will see (as already illustrated above in (3)) that *te* is very often displaced in verb clusters that do not deviate in word order from the standard Dutch 1-2-3-order. In other words, *te*-displacement seems to be a different phenomenon than German *zu*-displacement, and therefore also needs to receive a separate analysis (see also Salzmann (2019a)).

The outline of the case study is as follows. In chapter 4, I present the methodology and the results of the questionnaire study. In chapter 5, I present the formal analysis of the variation in presence and placement of *te* across the different cluster types as tested in the questionnaire. In chapter 6, I conclude the case study, and present a number of directions for future research.

[4] I have already presented and analysed a subpart of these data in Pots (2017).

CHAPTER 4

The questionnaire study

4.1 Introduction

In this chapter, I present the methodology and the results of the questionnaire study. In section 4.2, I present the methodology, in which the design, task and procedure of the questionnaire are discussed, followed by the presentation of the participants. In section 4.3, I present the results of the questionnaire study.

4.2 Methodology

4.2.1 Design

The questionnaire study was set up to investigate variation in the placement and presence of *te* in Dutch non-finite three verb clusters in the 1-2-3-order. I chose to limit the questionnaire to the 1-2-3-order in the cluster, because it would otherwise have been too long for the participants to fill out.

Four types of non-finite three verb clusters were tested in the questionnaire. In the first cluster type (1), henceforth ‘cluster type Ia’, the finite verb *zegt* ‘says’, which is in Verb Second position, selects a *te*-infinitive. Based on these selectional requirements, the first verb of the verb cluster, V1 *willen* ‘want’, should be a *te*-infinitive. The second cluster type, henceforth ‘cluster type Ib’ has the same structural configuration as cluster type Ia, but with different lexical items. The use of two sentences with the same structural configuration but different lexical items allows me to test whether the results of the questionnaire hold independently of specific lexical items. In the second cluster type

(3), henceforth 'cluster type II', the highest verb in the cluster, V1 *hoeven* 'need', selects a *te*-infinitive. V2 of the cluster, *gaan* 'go', should therefore be preceded by *te*. In the third cluster type (4), henceforth 'cluster type III', the second verb in the cluster, V2 *zitten* 'sit', selects the *te*-infinitive.[1] The third verb of the cluster, V3 *wachten* 'wait' should therefore be a *te*-infinitive. In all examples of these cluster types, the verb that selects the *te*-infinitive is given in bold, and the entire cluster is enclosed in square brackets.

(1) Anne **zegt** op haar comfortabele stoel [*te* willen$_1$ blijven$_2$ zitten$_3$].
Anne says on her comfortable chair to want.INF remain.INF sit.INF
'Anne says she wants to remain seated on her comfortable chair.'
CLUSTER TYPE IA

(2) Stijn **beweert** met het geld van zijn erfenis die grote villa [*te* hebben$_1$ kunnen$_2$ kopen$_3$].
Stijn claims with the money of his inheritance that big villa to have can buy
'Stijn claims to have been able to buy that big villa with his inheritance money.'
CLUSTER TYPE IB

(3) Koen zal vanwege de winterstop vandaag niet [**hoeven$_1$** *te* gaan$_2$ voetballen$_3$].
Koen will because.of the winter.break today not need.INF to go.INF play.football.INF
'Because of the winter break, Koen won't have to go and play football today.'
CLUSTER TYPE II

(4) Peter zal vanwege de nieuwe dienstregeling binnenkort nog langer op de trein [moeten$_1$ **zitten$_2$** *te* wachten$_3$].
Peter will because.of the new train.schedule soon even longer on the train must.INF sit.INF to wait.INF
'Because of the new train schedule, Peter will soon have to wait even longer for the train.'
CLUSTER TYPE III

The goal of the questionnaire study was fourfold. First, to test whether *te* can be absent from the cluster, even though selectional requirements dictate it should be present. Second, to test wether *te* can occur on a different verb than the one required by selection. Third, to test whether *te* can occur twice, even though selectional requirements dictate it should occur only once. Last, to test whether *te* behaves differently with respect to these three factors depending on the structural position of the *te*-selecting verb. In order to answer these questions, seven different versions of the four different cluster types were included in the questionnaire: three versions with a single *te*, once on V1, once on V2, and

[1] The verb *zitten* is used as an aspectual verb here, indicating progressive or durative aspect, see section 2.3.3.

once on V3, a version without *te*, and three versions with double *te*. These seven versions can be schematically represented as follows:

1. *te*-V1-V2-V3
2. V1-*te*-V2-V3
3. V1-V2-*te*-V3
4. V1-V2-V3
5. *te*-V1-*te*-V2-V3
6. *te*-V1-V2-*te*-V3
7. V1-*te*-V2-*te*-V3

Given that there are four cluster types, and 7 versions, the questionnaire consisted of 28 test items. 25 filler items were added to the questionnaire, leading to a questionnaire of 53 sentences in total. The items of the questionnaire were presented in randomised order. Prior to the questionnaire, a practice round of 5 sentences of different degrees of grammaticality was added. The full list of items is given in Appendix A. The questionnaire was created in Limesurvey©.

4.2.2 Task and procedure

The task was a judgment task, in which the participants were asked to rate the sentences of the questionnaire based on a five point Likert scale. For each sentence, the participants were asked to say the sentence out loud, which was done to prompt them to rate the sentence as if it was spoken, rather than to rate a written sentence. Then, they were asked to answer the following question: *'Is this a possible sentence in the variety of Dutch spoken in your immediate environment?'*. Below the question the five point Likert scale was presented, with 5 marked as 'certainly', 1 as 'certainly not', and 4, 3 and 2 as in-between stages that were not explicitly marked. They could also answer 'I don't know' if they really did not have any judgment about a sentence. For each sentence, a comment box was provided, where participants could write comments on their judgements.

The participants were first presented with an instruction text. They were familiarised with the fact that native speakers often have a quite clear idea about what a possible sentence in their language is, and that they can even say so regarding the *structure* of a sentence. It was explained that they would be presented with a number of sentences, each of which they were (i) asked to say aloud, and (2) to judge with respect to the question introduced above. In the instruction text, 'immediate environment', the phrase that was used in the question participants had to answer, was defined as 'friends, family, city or town'. The full instruction text can be found in Appendix A. Prior to the actual questionnaire, all participants were given the same set of 5 practice sentences. This was done so as to give the participants the time to get used to the task, prior to them rating the first sentences of the actual questionnaire. After the participants had completed the questionnaire, they were asked to answer a number of questions to collect some background information. The full list of background questions can be found in Appendix A.

4.2.3 Participants

The participants were recruited via social media, personal networks, and via the online newsletter *De Taalpost*. 531 native speakers of Dutch completed the questionnaire, 459 of which were included for the analysis. 70 participants were excluded based on the fact that they had lived or were living abroad for more than 10 per cent of their lives. This was done to control for possible influence from other languages or sparse use of Dutch on the judgments of the participants. Two participants were excluded based on their inconsistent ratings of the filler items that were completely ungrammatical, which these participants rated with a 4 or a 5. The mean age of the participants was 53 years (*SD*=12.5, range=18-99). The gender division was 250 female, 209 male. 361 participants were born in the Netherlands, 95 in Belgium, 2 in Surinam and 1 on Curaçao.

In Figure 4.1, the locations of the participants are presented on a map of the Dutch speaking language area.[2] Note that it is not the case that one point represents only one speaker: there are many locations for which there were multiple informants.

4.3 Results questionnaire study

4.3.1 Introduction

In this section, the results of the questionnaire study are presented. In subsection 4.3.2, some prerequisites regarding the data presentation and data encoding are discussed. In subsection 4.3.3, the general findings of the questionnaire study are presented, namely that *te* can be (i) dropped, (ii) raised, (iii) lowered, and (iv) doubled. In subsection 4.3.4, the focus lies on the frequencies and optionality of these four phenomena. I show that there is a two-by-two divide in the optionality of the four phenomena: *te*-drop and *te*-raising, if accepted, can be optional or obligatory, whereas *te*-lowering and *te*-doubling, if accepted, are always optional. In addition, the geographical spread of these phenomena is discussed as well. In subsection 4.3.5, variation and optionality across the different types of clusters is examined. The main observation of this subsection is that the presence and placement of *te* is most optional in cluster type II, less so in cluster type III and the least in cluster types Ia and Ib. In subsection

[2] All maps in this thesis are created in R (R Core Team 2014) with the packages Rworldmap and Rworldmap-extra (South et al. 2016).

Figure 4.1: Locations of the participants (case study I)

4.3.6, exploratory statistical techniques (Correspondence Analysis, Hierarchical Clustering and Multiple Correspondence Analysis) are used (i) to further explore and examine patterns and tendencies in the data, and (ii) to distinguish groups of speakers that have the same type of grammar with respect to the presence and placement of *te*. In subsection 4.3.7, I present the patterns of morphosyntactic variation and optionality that should be explained in the formal analysis.

4.3.2 Prerequisites

Before presenting the results of the questionnaire study, two prerequisites need to be discussed. The first prerequisite concerns the fact that in the case of cluster type Ib, cluster type II and cluster type III, there were a number of speakers who rejected all versions of the cluster. 40 speakers rejected all versions of cluster type Ib, 62 did so for cluster type II and 64 for cluster type III. The test sentence for cluster type Ib is repeated in (5).

(5) Stijn **beweert** met het geld van zijn erfenis die grote villa [
Stijn claims with the money of his inheritance that big villa
te hebben$_1$ kunnen$_2$ kopen$_3$].
to have can buy
'Stijn claims to have been able to buy that big villa with his inheritance money.'

Unfortunately, not many speakers commented on their reasons for rejecting this sentence. Three of them did provide feedback, though. The first comment was that the speaker preferred the object of the sentence *die grote villa* 'that big villa' to appear to the left of the phrase *met het geld van zijn erfenis* 'with his inheritance money'. The second comment was that the speaker felt that the phrase *met het geld van zijn erfenis* 'with his inheritance money' sounded somewhat 'awkward', because according to him/her it sounded as if the subject of the sentence was buying something with the money of his *own* will, rather than with money that he had inherited from somebody else (a grandfather/mother or so). Both comments are thus unrelated to the verb cluster itself. The third comment *was* about the cluster: the speaker wrote that he/she would prefer a cluster with a past participle as lexical verb, and proposed *gekocht te kunnen hebben* 'bought to can have' as an alternative. Note that in that cluster the hierarchical relation between the perfective auxiliary and the modal has been inverted.

For cluster type II, repeated her in (6), there are two types of comments that were given more than once. The first was made by southern (mostly Flemish) speakers, who commented that they prefer to use the modal *moeten* 'must' rather than *hoeven*, as in (7).[3] The second comment that was made more than once is that speakers prefer the second verb in the verb cluster, V2 *gaan* 'go', to be absent altogether, resulting in a two-verb rather than three-verb cluster (8).

(6) Koen zal vanwege de winterstop vandaag niet [**hoeven**$_1$ *te*
Koen will because.of the winter.break today not need.INF to
gaan$_2$ voetballen$_3$].
go.INF play.football.INF
'Because of the winter break, Koen won't have to go and play football today.'

(7) Koen zal vanwege de winterstop vandaag niet [**moeten**$_1$ gaan$_2$
Koen will because.of the winter.break today not need.INF go.INF
voetballen$_3$].
play.football.INF

[3] Given that *moeten*, unlike *hoeven*, does not select a *te*-infinitive, it was not an option for me to use this verb in the questionnaire instead.

'Because of the winter break, Koen won't have to go and play football today.'

(8) Koen zal vanwege de winterstop vandaag niet [**hoeven**$_1$ *te* voetballen$_2$].
Koen will because.of the winter.break today not need.INF to play.football.INF
'Because of the winter break, Koen won't have to go and play football today.'

As for cluster type III, repeated here in (9), the most frequent comment was that speakers prefer to leave out the aspectual *zitten* 'sit' from this cluster, thus using the two-verb cluster *moeten wachten* 'must wait', as in (10).

(9) Peter zal vanwege de nieuwe dienstregeling binnenkort nog langer op de trein [moeten$_1$ **zitten**$_2$ *te* wachten$_3$].
Peter will because.of the new train.schedule soon even longer on the train must.INF sit.INF to wait.INF
'Because of the new train schedule, Peter will soon have to wait even longer for the train.'

(10) Peter zal vanwege de nieuwe dienstregeling binnenkort nog langer op de trein [moeten$_1$ wachten$_2$].
Peter will because.of the new train.schedule soon even longer on the train must.INF wait.INF
'Because of the new train schedule, Peter will soon have to wait even longer for the train.'

Removing the ratings of speakers who rejected all versions of a type of cluster from the data set, we end up with the numbers of speakers listed in Table 4.1. These are speaker who accepted at least one version of the relevant cluster type. From here onwards, these numbers are used as the total number of speakers for the various cluster types in the discussion of the data.

Type of cluster	number of speakers
Cluster type Ia	459
Cluster type Ib	419
Cluster type II	397
Cluster type III	395

Table 4.1: Number of speakers per cluster type

The map in Figure 4.2 shows in which locations at least one speaker rejected all versions of cluster type Ib, cluster type II and/or cluster type III.

The second prerequisite concerns the way in which the data of the questionnaire study are encoded and presented. Recall that the data were collected by

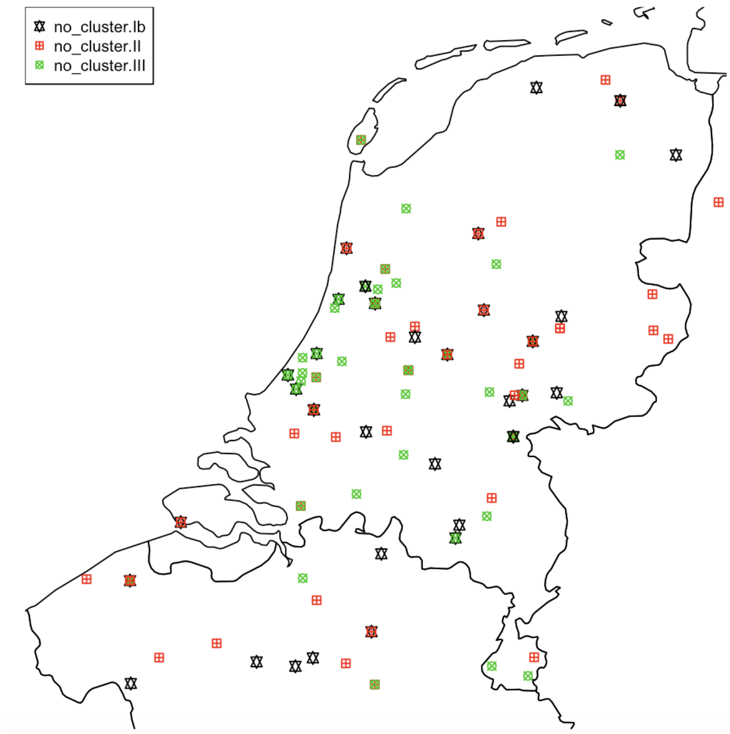


Figure 4.2: Locations in which all versions of cluster types Ib, II and/or III were rejected

asking native speakers to rate sentences using a 5-point Likert scale. However, to be able to make generalisations about whether a given phenomenon (for example *te*-drop) exists in a given cluster type or not, the data of the questionnaire study need to be encoded from ratings of 1 to 5, to 0 (= speaker finds this phenomenon in this cluster type ungrammatical) versus 1 (= speaker finds this phenomenon in this cluster type grammatical). Encoding the data in such as way gives us a tool to make generalisations and find clear patterns in the data. As an illustration, consider the difference between two data tables concerning *te*-drop in the four cluster types, i.e. the ratings of the test items of each cluster type in which *te* is absent. The first data table is one in which the data are presented using the 1 to 5 encoding corresponding to the ratings. This is Table 4.2 below. The second table is one in which the data are presented after all 1-3 ratings are encoded as 0 (i.e. 'No *te*-drop'), and all 4-5 ratings are encoded as 1 (i.e. '*Te*-drop'). This is Table 4.3.

Comparing the two tables, we see that the first type of table contains more detailed information of the exact spread of the total frequency over the five

Cluster	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5	Total
Type Ia	413 (90,0%)	29 (6,3%)	9 (2,0%)	4 (0,25%)	4 (0,25%)	459 (100%)
Type Ib	382 (91,2%)	0 (0,0%)	0 (0,0%)	27 (6,4%)	10 (2,4%)	419 (100%)
Type II	103 (25,9%)	64 (16,1%)	25 (6,3%)	100 (25,2%)	105 (26,5%)	397 (100%)
Type III	14 (3,5%)	3 (0,8%)	6 (1,6%)	62 (15,6%)	310 (78,5%)	395 (100%)

Table 4.2: Frequency overview of *te*-drop per cluster type version I

Cluster	No *te*-drop	*Te*-drop	Total
Type Ia	451 (98,3%)	8 (0,7%)	459 (100%)
Type Ib	382 (91,2%)	37 (8,8%)	419 (100%)
Type II	191 (48,1%)	208 (51,9%)	397 (100%)
Type III	20 (5,1%)	375 (94,9%)	395 (100%)

Table 4.3: Frequency overview of *te*-drop per cluster type version II

rating options, whereas the second type of table gives us directly a clear picture of the frequency distribution between speakers who do not accept *te*-drop in a given cluster type and speakers who do. As I find it more important to make generalisations about the existence of a given phenomenon in a cluster type than to give the reader a very detailed picture of the frequency spread over the five rating options, I choose to present the data using the 0 versus 1 encoding rather than using the 1-5 encoding. Furthermore, I also want to investigate frequencies of intra-speaker optionality of the phenomena. For example, I want to know how many speakers optionally drop *te* in the four cluster types, and compare that frequency with the of speakers who need *te* to be present and the frequency of speakers who need *te* to be absent. Such a table is given for *te*-drop in Table 4.4. Speakers who only accept the cluster type with *te* present are given in the first column ('No *te*-drop'), speakers who accept the cluster type both with and without *te* are given in the second column ('Opt. *te*-drop'), and speakers who only accept the cluster type without *te* are given in the third column ('Obl. *te*-drop'). Such a data table is impossible to make if we do not use the encoding of 0 versus 1.

Cluster	No *te*-drop	Opt. *te*-drop	Obl. *te*-drop	Total
Type Ia	451 (98,3%)	8 (0,7%)	0 (0%)	459 (100%)
Type Ib	382 (91,2%)	35 (8,4%)	2 (0,4%)	419 (100%)
Type II	191 (48,1%)	187 (47,1%)	19 (4,8%)	397 (100%)
Type III	20 (5,1%)	152 (38,5%)	223 (56,4%)	395 (100%)

Table 4.4: Frequency overview of *te*-drop per cluster type

All encoding of the data is done as follows: all 1-3 ratings are encoded as 0 (= ungrammatical), and all 4-5 ratings are encoded as 1 (= grammatical).[4] This encoding is used in the presentation of the frequency and optionality patterns in subsection 4.3.4 and in the presentation of the variation and optionality across cluster types in subsection 4.3.5. It is furthermore used in subsection 4.3.6 as input for the exploratory statistical techniques Correspondence Analysis and Hierarchical Clustering. However, in the last part of this latter subsection (4.3.6.4), I return to the 1-5 encoding, and use this as the input for the exploratory statistical technique Multiple Correspondence Analysis.

4.3.3 General findings

The general findings of the first questionnaire study are that (i) *te* can be dropped, i.e. be absent from the cluster altogether, (ii) *te* can be raised, i.e. appear on a higher verb than is required by selectional requirements, (iii) *te* can be lowered, i.e. appear on a lower verb than is required by selectional requirements, and (iv) *te* can be doubled, i.e. appear twice, even though selectional requirements dictate that it should occur only once. For each phenomenon an example is given below, using cluster type II for illustration.[5]

(11) Koen zal vanwege de winterstop vandaag niet [**hoeven**$_1$ gaan$_2$ voetballen$_3$].
Koen will because.of the winter.break today not need.INF go.INF play.football.INF
'Because of the winter break, Koen won't have to go and play football today.' *te*-DROP

(12) Koen zal vanwege de winterstop vandaag niet [*te* **hoeven**$_1$ gaan$_2$ voetballen$_3$].
Koen will because.of the winter.break today not to need.INF go.INF play.football.INF
'Because of the winter break, Koen won't have to go and play football today.' *te*-RAISING

[4] Note that a judgment '3' is not to be taken as a sign that the speaker was unsure how to rate the sentence, given that s/he always had the option of assigning 'I don't know'. If a speaker assigned a '3' to a sentence, this is therefore taken to mean that s/he finds the sentence somewhat marginal, but not completely. I prefer to combine this with '1' and '2' as ungrammatical to make sure not to overgeneralise the grammaticality of the phenomena discussed in this study.

[5] Note that there is only one version of *te*-doubling given as illustration. The possible doubling patterns for all cluster types are reported in detail in subsection 4.3.4.

(13) Koen zal vanwege de winterstop vandaag niet [**hoeven**$_1$ gaan$_2$
Koen will because.of the winter.break today not need.INF go.INF
te voetballen$_3$].
to play.football.INF
'Because of the winter break, Koen won't have to go and play football today.' *te*-LOWERING

(14) Koen zal vanwege de winterstop vandaag niet [**hoeven**$_1$ *te*
Koen will because.of the winter.break today not need.INF to
gaan$_2$ *te* voetballen$_3$].
go.INF to play.football.INF
'Because of the winter break, Koen won't have to go and play football today.' *te*-DOUBLING

Let us now look at the frequencies and optionality of each of these phenomena in turn.

4.3.4 The frequencies and optionality of the phenomena

Let us begin with the frequencies of optional and obligatory *te*-drop. In Table 4.5, the number of speakers are given for each cluster type who do not accept the version of the sentence in which *te* is absent ('No *te*-drop'), those who accept both the version in which *te* is absent and in which it is present ('Opt. (optional) *te*-drop'), and those who only accept the version in which *te* is absent ('Obl. (obligatory) *te*-drop').

Cluster	No *te*-drop	Opt. *te*-drop	Obl. *te*-drop	Total
Type Ia	451 (98,3%)	8 (0,7%)	0 (0%)	459 (100%)
Type Ib	382 (91,2%)	35 (8,4%)	2 (0,4%)	419 (100%)
Type II	191 (48,1%)	187 (47,1%)	19 (4,8%)	397 (100%)
Type III	20 (5,1%)	152 (38,5%)	223 (56,4%)	395 (100%)

Table 4.5: Frequency overview of *te*-drop per cluster type

This table shows that in cluster types Ia and Ib, virtually all speakers need *te* to be present, even though there is a small group of speakers who optionally allow *te*-drop in cluster type Ib, and a handful who do so for cluster type Ia. In cluster type II, there is a much smaller group of people who require *te* to be present, a similarly sized group of speakers for whom *te* is optional, and a very small group of speakers who need *te* to be absent. In cluster type III, only a very small group of speakers need *te* to be present, while for a large group of speakers *te* is optional, and the largest group needs *te* to be absent. Given that large groups of speakers allow *te*-drop in cluster types II and III, it would be interesting to see the geographical distribution of speakers who do not allow *te*-drop in these clusters. It is furthermore interesting to see whether the geographical spread is different for the two cluster types. This is visualised in Figure 4.3. As can be seen in the map, the speakers who do not allow *te*-drop with *hoeven* 'need' (cluster type II) are spread more or less evenly over the entire language area, though only to a limited extent in the northern part of the Netherlands. We have to keep in mind, though, that there were also slightly less data points for the northern part of the Netherlands to begin with (see also Figure 4.1).

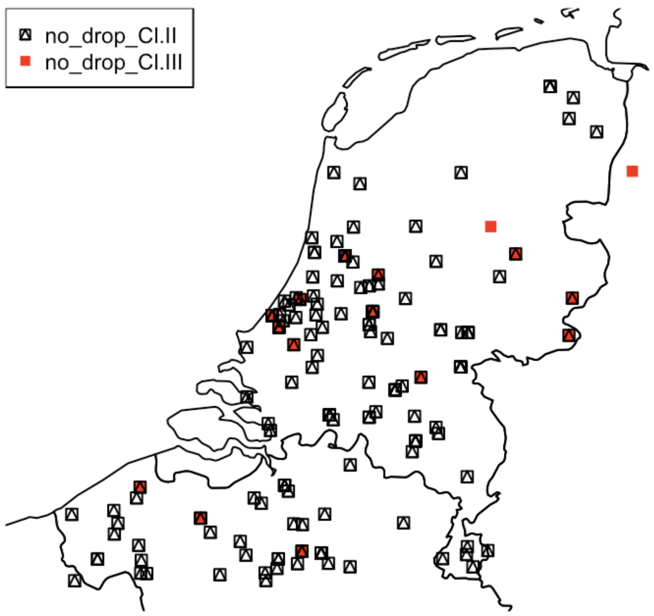


Figure 4.3: Geographical spread of unacceptability of *te*-drop in cluster types II and III

Let us now move on to the optionality frequencies of the second phenomenon: *te*-raising. As *te* is already in the highest position of the cluster in cluster types Ia and Ib (namely on V1), *te*-raising is not a possible phenomenon in these two cluster types. Therefore, only the frequencies for cluster types II and III are presented in Table 4.6.

Table 4.6 shows the numbers of speakers who only accept *te* in the position as by selectional requirements ('No *te*-raising'), those who accept *te* both in the correct position and in a raised position ('Optional *te*-raising'), and those

Cluster	No *te*-raising	Opt. *te*-raising	Obl. *te*-raising	Total with *te*
Type II	193 (51,1%)	149 (39,4%)	36 (9,5%)	378 (100%)
Type III	132 (76,8%)	31 (18,0%)	9 (5,2%)	172 (100%)

Table 4.6: Frequency overview of *te*-raising per cluster type

who only accept *te* to be raised ('Obligatory *te*-raising').[6] The frequencies show that *te*-raising can be either optional or obligatory for both cluster type II and cluster type III. Furthermore, we see that *te*-raising is much more frequent in cluster type II (39,4% optional *te*-raising and 9,5% obligatory *te*-raising) than in cluster type III (18,0% optional *te*-raising and 5,2% obligatory *te*-raising). Recall that in cluster type II, *te* should appear on V2. This means that if *te* raises, it can only raise to one position, namely onto V1. In cluster type III, however, *te* should appear on V3, which means that there are two possible positions for it to raise to in this cluster type: V1 and V2. In Table 4.7, the frequencies of each option is given, and additionally also the frequencies of speakers who allow raising to both V1 and V2.

Cluster	Raising to V1 and V2	Only to V1	Only to V2	Total
Type III	6 (15,0%)	15 (37,5%)	19 (47,5%)	40 (100%)

Table 4.7: Frequency overview of position of *te*-raising in cluster types III

Table 4.7 shows that the lowest percentage of *te*-raisers in cluster type III allow raising to both V1 and V2. The biggest group only raises to the immediately higher (namely onto V2, 47,5%), and a slightly smaller group only raises to V1 (37,5%). Let us know look at the geographical spread of *te*-raising. Three things are of particular interest here. First, we want to see whether the phenomenon of *te*-raising is confined to a specific part of the language area. Second, we want to find out whether *te*-raising has a different geographical distribution for cluster type II compared to cluster type III: in other words, we want to know whether *te* behaves differently when selected by *hoeven* compared to *zitten*. Third, we want to see if there is any specific divide in geographical patterns between obligatory and optional *te*-raising. That is, we want to see whether obligatory *te*-raising is specific to a given part of the language area. The geographical distribution of *te*-raising is given in Figure 4.4.

The map shows that *te*-raising is a phenomenon that is common all over the language area, but again, as we saw for *te*-drop in Figure 4.3, the least frequent in the northern part of the Netherlands. Again, we have to keep in

[6]Note that the column 'Total with *te*' gives the numbers of speakers who accept at least one of the versions in which *te* occurs, i.e. speakers with obligatory *te*-drop are subtracted from the total number of speakers per cluster type.

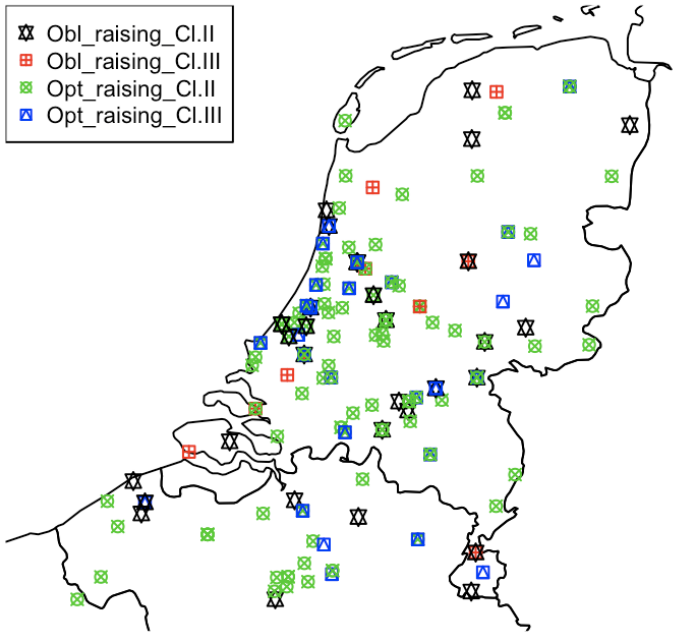


Figure 4.4: Geographical spread of *te*-raising

mind that there are slightly less data points for this part of the Netherlands in this study in general. What is clear from the map, though, is that *te*-raising in cluster type III, be it optional (given in blue) or obligatory (given in red), is hardly attested in the north. Note also that many of the locations only allow *te*-raising in cluster type II or in cluster type III. This means that the underlying mechanism for *te*-raising cannot be a simple parametric switch-on or switch-off that applies to all instances of *te*-raising. I return to this hypothesis below.

We move on to the next phenomenon: *te*-lowering. Recall that in cluster type III, *te* is already on the lowest verb of the cluster (V3), which means that *te*-lowering is not a possible phenomenon for this cluster type. Therefore, in Table 4.8, only the frequencies of *te*-lowering for cluster types Ia, Ib and II are presented.

Table 4.8 gives the numbers of speakers who only accept *te* in the correct position ('No *te*-lowering'), those who accept *te* both in the correct position and in a lowered position ('Opt. *te*-lowering'), and those who only allow *te*-lowering ('Obl. *te*-lowering'). It turns out that *te*-lowering is always optional: none of the

Cluster	No *te*-lowering	Opt. *te*-lowering	Obl. *te*-lowering	Total with *te*
Type Ia	403 (87,8%)	56 (12,2%)	0 (0%)	459 (100%)
Type Ib	356 (85,0%)	63 (15,0%)	0 (0%)	419 (100%)
Type II	338 (89,4%)	40 (10,6%)	0 (0%)	378 (100%)

Table 4.8: Frequency overview *te*-lowering per cluster type

speakers only accepts the version with a lowered *te*. This sets *te*-lowering apart from the other two phenomena discussed so far; both *te*-drop and *te*-raising are the only option for some of the speakers. A second difference between *te*-lowering on the one hand and *te*-drop and *te*-raising on the other, is that the frequencies for *te*-lowering are very comparable across cluster types (12,2% for cluster type Ia, 15,0% for cluster type Ib and 10,6% for cluster type II), whereas *te*-drop is much more frequent in cluster type II (51,9%) compared to cluster types Ia (0,7%) and Ib. (8,8%), and even more frequent in cluster type III (94,9%). Similarly, *te*-raising is much more frequent in cluster type II (48,9%) than in cluster type III (23,2%).[7] Both these facts might be an indication that the mechanism behind *te*-lowering is very different from the one behind *te*-drop or *te*-raising. I will return to this hypothesis later in subsection 4.3.6.

Recall that in cluster types Ia and Ib, *te* should appear on V1. Since there are three verbs in the clusters under consideration, *te* has two potential positions to lower to: V2 and V3. In the following table, the frequencies of lowering to both V2 and V3 are given, as well as frequencies of lowering only to V2, and only to V3.

Cluster	Lowering to V2 and V3	Only to V2	Only to V3	Total
Type Ia	3 (4,8%)	51 (80,9%)	9 (14,3%)	63 (100%)
Type Ib	10 (15,2%)	49 (74,2%)	7 (10,6%)	66 (100%)

Table 4.9: Frequency overview of subtypes of *te*-lowering in cluster types Ia and Ib

Table 4.9 shows that the most frequent pattern is that speakers only lower to V2, namely 80,9% of the speakers in the case of cluster type Ia and 74,2% of the speakers in the case of cluster type Ib. Very few speakers allow lowering to both V2 and V3 (4,8% in cluster type Ia and 15,2% in cluster type Ib), and likewise, very few speakers allow lowering only to V3 (14,3% for cluster type Ia and 10,6% for cluster type Ib).

Let us now look at the geographical distribution of *te*-lowering, presented in Figure 4.5. As can be seen from the map, *te*-lowering is a phenomenon that

[7]The frequencies given here are percentages of optional and obligatory *te*-drop combined, and of optional and obligatory *te*-raising combined.

is spread over the entire language area. We furthermore see that it is not the case that in each location in which *te*-lowering is allowed, it is allowed for all cluster types.

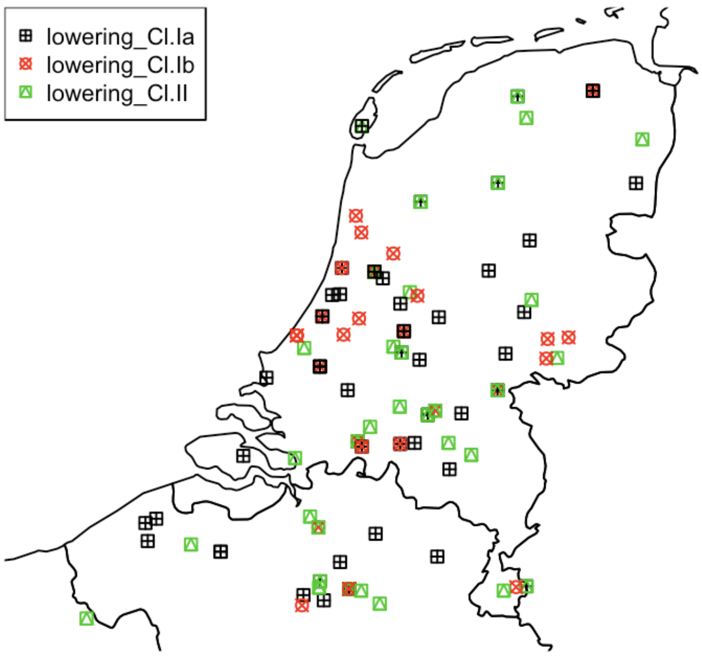


Figure 4.5: Geographical spread of *te*-lowering

Finally, we look at the optionality frequencies of the fourth phenomenon: *te*-doubling. In Table 4.10, the frequency patterns for *te*-doubling are presented.

Cluster	No *te*-doubling	Opt. *te*-doubling	Obl.*te*-doubling	Total with *te*
Type Ia	405 (88,2%)	54 (11,8%)	0 (0%)	459 (100%)
Type Ib	375 (89,5%)	44 (10,5%)	0 (0%)	419 (100%)
Type II	317 (83,9%)	61 (16,1%)	0 (0%)	378 (100%)
Type III	160 (93,0%)	12 (7,0%)	0 (0%)	172 (100%)

Table 4.10: Frequency overview *te*-doubling per cluster type

In Table 4.10, the frequencies are given of the numbers of speakers who do not allow *te*-doubling ('No *te*-doubling'), those who allow both a version with *te*-doubling and a version without ('Optional *te*-doubling'), and those who only allow a version of the cluster type in which *te* is doubled ('Obligatory *te*-doubling'). As can be seen from the table, there are no speakers for whom *te*-doubling is obligatory. In this respect, *te*-doubling patterns together with *te*-lowering, and not with *te*-raising and *te*-drop. Also like *te*-lowering and unlike the other two phenomena, *te*-doubling shows more or less the same frequencies across the four cluster types, even though it is considerably less frequent in cluster type III in absolute numbers.

Recall from subsection 4.2.1 that all three theoretically possible configurations of *te*-doubling were tested for each cluster type: doubling on V1 and V2 (*te*-V1-*te*-V2-V3), doubling on V1 and V3 (*te*-V1-V2-*te*-V3) and doubling on V2 and V3 (V1-*te*-V2-*te*-V3). In Table 4.11, the frequencies for each configuration are given for all cluster types. Note that there are speakers who allow more than one type of doubling configuration. The numbers given in the table are the numbers of speakers who accept that specific configuration, but they do not necessarily accept *only* that configuration. In other words, the column names above the presented frequencies should be read as 'No *te*-doubling in this configuration'; 'Optional *te*-doubling in this configuration' and 'Obligatory *te*-doubling in this configuration'.

Cluster	Configuration	No *te*-doubling	Opt.	Obl.	Total with *te*
Type Ia	*te*-V1-*te*-V2-V3	442 (96,3%)	17 (4,7%)	0 (0,0%)	459 (100%)
	te-V1-V2-*te*-V3	446 (97,2%)	13 (2,8%)	0 (0,0%)	459 (100%)
	V1-*te*-V2-*te*-V3	422 (91,9%)	37 (8,1%)	0 (0,0%)	459 (100%)
Type Ib	*te*-V1-*te*-V2-V3	411 (98,1%)	8 (1,9%)	0 (0,0%)	419 (100%)
	te-V1-V2-*te*-V3	410 (97,9%)	9 (2,1%)	0 (0,0%)	419 (100%)
	V1-*te*-V2-*te*-V3	388 (92,6%)	31 (7,4%)	0 (0,0%)	419 (100%)
Type II	*te*-V1-*te*-V2-V3	340 (89,9%)	38 (10,1%)	0 (0,0%)	378 (100%)
	te-V1-V2-*te*-V3	345 (91,3%)	33 (8,7%)	0(0,0%)	378 (100%)
	V1-*te*-V2-*te*-V3	366 (96,8%)	12 (3,2%)	0 (0,0%)	378 (100%)
Type III	*te*-V1-*te*-V2-V3	167 (97,1%)	5 (2,9%)	0 (0,0%)	172 (100%)
	te-V1-V2-*te*-V3	168 (97,7%)	4 (2,3%)	0 (0,0%)	172 (100%)
	V1-*te*-V2-*te*-V3	164 (95,3%)	8 (4,7%)	0 (0,0%)	172 (100%)

Table 4.11: Frequency overview of *te*-doubling per configuration and type of cluster

The following observations can be made based on the frequencies presented in Table 4.11. First, in both cluster types Ia and Ib, the most frequent configuration is in fact the one in which neither of the two *te*'s occurs on the verb which selectional requirements dictate it should occur on. That is, *te* should occur on V1, but the most frequent option is 'V1-*te*-V2-*te*-V3'. Second, in cluster type II the most frequent doubling configuration is the one in which *te* occurs on the verb it should occur on (V2), and one verb higher (V1) (i.e. *te*-V1-*te*-V2-V3), but the configuration in which neither one of the two *te*'s occurs on the correct verb is also quite frequent (i.e. *te*-V1-V2-*te*-V3). Third, in cluster type III the most frequent option is, like in cluster type II, the configuration in which one *te* occurs on the verb it should occur on (V3) and the other on the verb that is one position higher (V2): V1-*te*-V2-*te*-V3. Let us know look at the geographical distribution of *te*-doubling. The corresponding map is given in Figure 4.6. As was the case with the other phenomena discussed above, *te*-doubling is widespread and not specific to one part of the language area. In contrast to the other phenomena, however, *te*-doubling in different cluster types seems occur more often in the same location than not. This might be an indication that the underlying mechanism for *te*-doubling is the same across cluster types. I return to this hypothesis below.

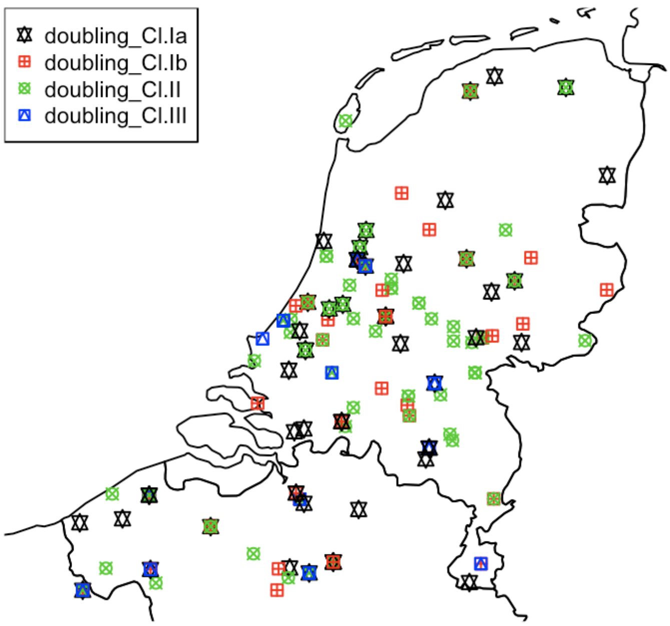


Figure 4.6: Geographical spread of *te*-doubling

In summary, when focussing on the optionality and obligatoriness of the four phenomena, we have seen the following. First, *te*-drop and *te*-raising look similar in the sense that among the speakers who allow *te*-drop and *te*-raising, these phenomena are either optional or obligatory. In contrast, *te*-lowering and *te*-doubling are, if accepted, always optional. Second, *te*-drop and *te*-raising also show the same pattern in the sense that the frequencies across cluster types are not comparable: *te*-drop is least frequent in cluster types Ia and Ib, much more frequent in cluster type II and most frequent in cluster type III, whereas *te*-raising is much more frequent in cluster type II compared to cluster type III. Again in contrast, this does not seem to hold for the other two phenomena: *te*-lowering and *te*-doubling have comparable frequencies across cluster types Ia, Ib and II.[8] These findings lead to the hypothesis that *te*-lowering and *te*-doubling should be analysed differently from *te*-raising and *te*-drop. This hypothesis will be further explored and tested against the data in subsections 1.3.6 and 1.3.7, by means of the exploratory statistical techniques Correspondence Analysis and Hierarchical Clustering. Before we move on to that, however, let us first have a closer look at the optionality of the four phenomena, by considering their weighted frequencies, and the number of options that speakers allow per cluster type. This is done in the next subsection.

4.3.5 Variation and optionality across cluster types

In this subsection, we look at the weighted frequencies of the four phenomena. Weighted frequencies give a better and more direct overview of how frequent the phenomena are compared to one other for each cluster type, as it takes the optionality of the phenomena into account in calculating the frequencies. Recall from the subsection 4.2.1 that each cluster type was presented to the speakers in seven different versions, repeated here below for convenience:

1. *te*-V1-V2-V3
2. V1-*te*-V2-V3
3. V1-V2-*te*-V3
4. V1-V2-V3
5. *te*-V1-*te*-V2-V3
6. *te*-V1-V2-*te*-V3
7. V1-*te*-V2-*te*-V3

For each cluster type, one of the first three versions is the one in which *te* occurs on the verb on which it should occur ('*te* in situ'), and the other two versions are either cases of *te*-raising or *te*-lowering. The fourth version is the one in which *te* is dropped. The last three are cases of *te*-doubling.

To calculate the weighted frequencies of the four phenomena and of *te* in situ, I checked for each speaker how many out of the seven versions they accept for that cluster type. Per speaker and per cluster type, 1 point is divided over the number of options the speaker allows for that specific cluster type. For example, if a speaker allows only one version for cluster type Ia, that version gets 1 point. However, if a speaker allows two versions for cluster type Ia, both versions get 0,5 point. If a speaker allows three versions, all three versions get 0,33 point, etc. This is done for all four cluster types.

[8] *Te*-lowering is not possible in cluster type III given that *te* is already on the lowest verb of the cluster based on selectional requirements; *te*-doubling is less frequent in cluster type III compared to the other cluster types.

After having distributed the points over the accepted versions for all speakers for all cluster types, one extra step was taken in order to get the weighted frequencies of the phenomena. Recall, for example, that there are two possible *te*-lowering versions for cluster types Ia and Ib: V1-*te*-V2-V3 and V1-V2-*te*-V3. In other to get the weighted frequencies of *te*-lowering for these two cluster types, we thus have to cumulate the frequencies for both versions. Similarly, there are two possible *te*-raising versions for cluster type III: *te*-V1-V2-V3 and V1-*te*-V2-V3, so these frequencies are cumulated as well. Finally, for all cluster types, there are three possible doubling configurations: *te*-V1-*te*-V2-V3, *te*-V1-V2-*te*-V3 and V1-*te*-V2-*te*-V3. For each cluster type, the frequencies of the three doubling versions are cumulated as well. The weighted frequencies of all four phenomena and of *te* in situ are presented in Table 4.12.

Cluster	*Te* in situ	*Te*-raising	*Te*-lowering	*Te*-drop	*Te*-doubling	Total
Type Ia	399 (86,9%)		30,7 (6,7%)	2,5 (0,6%)	26,8 (5,8%)	459 (100%)
Type Ib	348,3 (83,1%)		30,2 (7,2%)	18,1 (4,3%)	22,4 (5,4%)	419 (100%)
Type II	185,8 (46,8%)	76,8 (19,3%)	17,3 (4,4%)	88,9 (22,4%)	28,2 (7,1%)	397 (100%)
Type III	77,1 (19,6%)	30,4 (7,7%)		280,6 (71,0%)	6,9 (1,7%)	395 (100%)

Table 4.12: Weighted frequencies of all phenomena per type of cluster

The following observations can be made based on Table 4.12. First, the weighted frequencies of all phenomena are very similar for cluster types Ia and Ib. *Te*-drop is slightly more frequent in cluster type Ib, but besides that the weighted frequencies are almost the same. This means that we can generalise the possible behaviour of *te* in a non-finite three-verb cluster when selected by a verb in Verb Second position beyond the two specific test sentences that were used in this questionnaire study. Second, as was already observed in the previous section based on the normal frequencies, the weighted frequencies also show a big difference in *te*-raising and *te*-drop across cluster types, whereas this is not the case for *te*-lowering and *te*-doubling (again, with the exception of *te*-doubling in cluster type III, which is very infrequent). Third, we see that the frequencies are spread out most over all phenomena in cluster type II, followed by cluster type III, and much less so in cluster types Ia and Ib. For the latter two cluster types, *te* in situ is by far the most frequent option (with a weight of 86,9% and 83,1% respectively), with no other phenomenon having a weighted frequency higher than around 7%. As for cluster type III, *te*-drop is the most frequent option (71,0%), but *te* in situ also reaches a noticeable percentage (19,6%). In the case of cluster type II, there are two phenomena besides *te* in situ (46,8%) that are quite frequent: *te*-raising and *te*-drop. In other words, cluster types Ia and Ib show the least amount of variation and optionality, and in cluster type II we find the most variation and optionality. Cluster type III occupies an intermediate position.

In order to visualise this degree of variation and optionality per cluster type in a different way, we can investigate the number of versions that are accepted for each cluster type. That is, we can calculate for each cluster type the number of speakers that allow 0 versions for a given cluster type, those that allow 1 version, 2 versions, etc. These numbers are presented in Table 4.13, and they are also visualised (up until 5 options) in the stacked bar chart in Figure 4.7.

Type of cluster	0 versions	1 version	2 versions	3 versions	4 versions	5 versions	6 versions	7 versions
Cluster type Ia	0 (0%)	352 (76,7%)	72 (15,7%)	31 (6,8%)	2 (0,4%)	0 (0%)	2 (0,4%)	0 (0%)
Cluster type Ib	40 (8,7%)	312 (68,0%)	80 (17,4%)	23 (5,0%)	4 (0,9%)	0 (0%)	0 (0%)	0 (0%)
Cluster type II	62 (13,5%)	127 (27,7%)	139 (30,3%)	96 (20,9%)	29 (6,4%)	4 (0,8%)	2 (0,4%)	0 (0%)
Cluster type III	64 (13,9%)	230 (50,1%)	125 (27,3%)	19 (4,1%)	6 (1,3%)	15 (3,3%)	0 (0%)	0 (0%)

Table 4.13: Frequencies of numbers of versions per cluster type

As expected from the results so far, the most frequent option for cluster types Ia and Ib is the one in which speakers allow only one version of that cluster: 76,7% and 68,0% respectively. This is also the case for cluster type III, even though the percentage is lower than in cluster types Ia and Ib, namely 50,1%. For cluster type II, we see that it is most frequent for speakers to accept two versions (30,3%), followed by 1 version (27,7%), and after that, three versions (20,9%). Also in the other cluster types, there are speakers who accept three versions, but much less so than in cluster type II. For all cluster types, there are speakers who allow four versions, which is again the most frequent for cluster type II. Five accepted versions is only attested for clusters type II and III (slightly more so for cluster type III). Finally, there two speakers that accept six versions of cluster type Ia, and two that accept six versions of cluster type II. There are no speakers who allow all seven versions for any of the cluster types. Summing up, we can see that it is most common for cluster type II for speakers to accept more than two versions of the cluster, followed by cluster type III. In other words, we find most variation and optionality with cluster type II, after that in cluster type III, and the least in cluster types Ia and Ib.

4.3.6 Statistical analysis of the patterns

4.3.6.1 Correspondence Analysis

In this subsection, I use Correspondence Analysis (henceforth CA), an exploratory statistical technique, to further explore the patterns and tendencies in the questionnaire data.[9] Recall from subsection 4.3.4 that when looking at the optionality of the four phenomena, we found that *te*-drop and *te*-raising can be optional or obligatory, while *te*-lowering and *te*-doubling, if accepted, are always optional. These findings lead to the hypothesis that *te*-lowering and *te*-doubling should be analysed differently from *te*-drop and *te*-raising. One way to investigate this hypothesis is by means of exploratory statistical techniques, like CA, which can be used to explore and visualise patterns in multivariate

[9]The calculations presented and discussed in this subsection were carried in R (R Core Team 2014) out using the CA function of the FactoMineR package Husson et al. (2014).

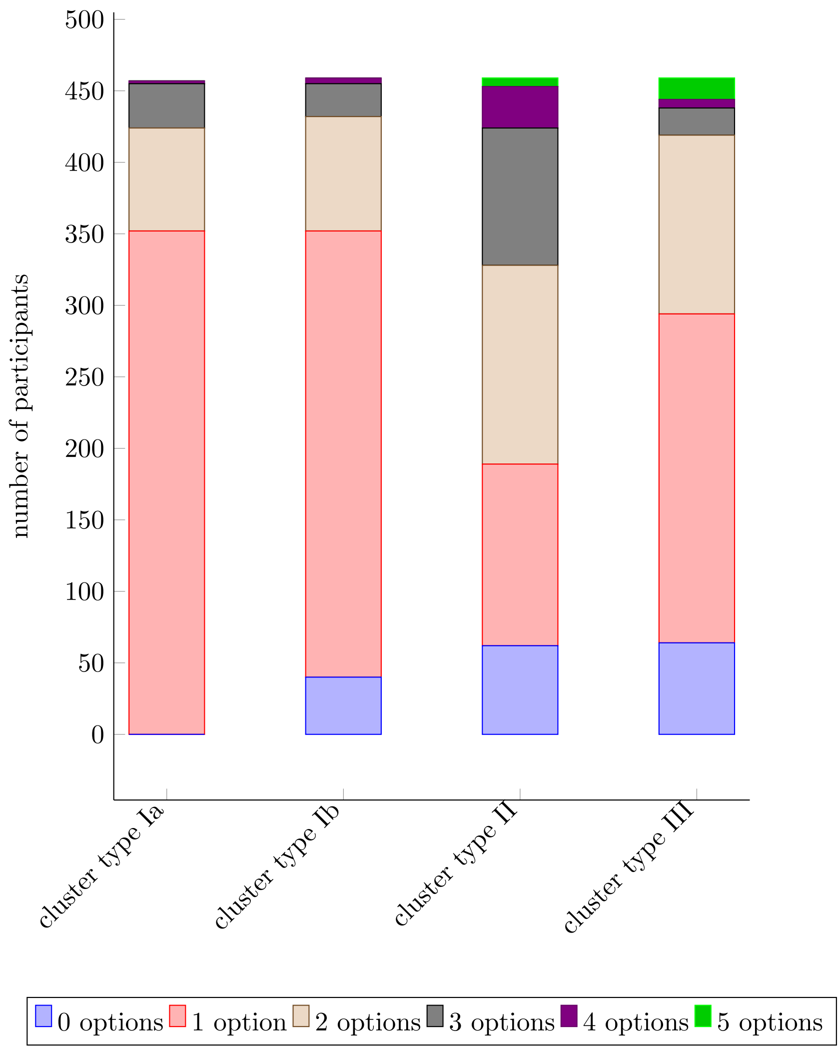


Figure 4.7: Frequencies of numbers of versions per cluster type (stacked bar chart)

data (Greenacre 2007, Levshina 2015). That is, it can be used to identify systematic relationships between variables (the phenomena per cluster type in the case of this study), which is what is needed to further test the hypothesis that *te*-raising and *te*-drop behave differently from *te*-lowering and *te*-doubling. In order to prepare the questionnaire data for the CA, the raw data table had to be adjusted. The raw data table has the participants as rows. The cells contained the rating a specific participant gave to a specific test sentence. As an illustration, the upper lefthand corner of the raw data table is given in Table 4.14.

	testitem1 *situ.Ia*	testitem2 *drop.Ia*	testitem3 *lower-to-V2.Ia*	testitem4 *lower-to-V3.Ia*	testitem5 *double-on-V1-V2.Ia*	...
PP1	5	1	1	1	1	...
PP2	5	1	2	2	1	...
PP3	5	2	2	1	1	...
PP4	5	2	1	1	1	...
PP5	5	2	1	1	1	...
PP6	5	1	4	1	5	...
PP7	5	1	2	2	1	...
...	...	...	...	...	...	...

Table 4.14: Upper lefthand corner of raw data table

To prepare the raw data table for CA, the following three steps were taken. As already discussed in subsection 4.3.2, all the 1's, 2's and 3's were replaced by a '0' (i.e. ungrammatical for the speaker), and the 4's and 5's were replaced by a '1' (i.e. grammatical for the speaker). Second, the test items were turned into 'phenomena' per cluster type. This was done in order to investigate the *general* relationships *te*-in-situ, *te*-drop, *te*-raising, *te*-lowering and *te*-doubling across the four cluster types. That is, differences among *te*-raising to V2 *te*-raising to V1, among *te*-lowering to V2 and *te*-lowering to V3, and among the three *te*-doubling configurations are thus abstracted away from. A more fine-grained, but therefore less generalisable, investigation in which the interrelation among the different *te*-raising, *te*-lowering and *te*-doubling configurations is included, is done in 4.3.6.4.

The test items were turned in to 'phenomena' as follows. The two test items in which *te*-lowering was tested for cluster type Ia and Ib (lowering to V1 and lowering to V2) were collapsed together to get the phenomenon '*te*-lowering in cluster type Ia' and '*te*-lowering in cluster type Ib' respectively. Similarly, all three versions of *te*-doubling (doubling on V1 and V2, doubling on V1 and V3, and doubling on V2 and V3) were collapsed together for each cluster type. For cluster type III, both raising to V2 and raising to V1 were collapsed together as '*te*-raising in cluster type III'. This leaves us with 17 different phenomena.[10] Finally, the data table is inverted, so that the phenomena are rows, and the participants columns. The upper lefthand corner of the data table used for CA is given in Table 4.15.

CA consists of two steps. In the first step, the raw data table (in this case, the complete version of Table 4.15) is converted into a so-called distance matrix. In order to get a distance matrix, the 'distance' between every pair of

[10]Four phenomena for cluster types Ia and Ib: *te*-in-situ, *te*-drop, *te*-lowering and *te*-doubling, five phenomena for cluster type II: *te*-in-situ, *te*-drop, *te*-lowering and *te*-doubling, and four for cluster type III: *te*-in-situ, *te*-drop, *te*-raising and *te*-doubling.

	PP1	PP2	PP3	PP4	PP5	PP6	PP7	...
situ.Ia	1	1	1	0	0	1	1	...
drop.Ia	0	0	0	0	0	1	0	...
lower.Ia	0	0	0	0	0	0	0	...
double.Ia	0	0	0	0	0	1	0	...
situ.Ib	1	0	1	1	1	1	1	...
...	...	...	...	...	...	...	...	...

Table 4.15: Upper lefthand corner of raw data table as input for CA

phenomena is calculated. The more the same speakers accept both phenomena, the smaller the distance between them is. The distance matrix of this study is a matrix of 17x17 (recall that we have 17 phenomena), with the phenomena both as rows and as columns. As comparing phenomenon X with phenomenon Y results in the same distance as comparing phenomenon Y with phenomenon X, the distance matrix is symmetrical across its diagonal, and we only need to look at one half of the table. As an illustration, the upper lefthand corner of the distance matrix is given in Table 4.16.

	situ.Ib	lower.Ib	drop.Ib	double.Ib	situ.Ia	lower.Ia	drop.Ia	...
situ.Ib	0.0							
lower.Ib	18.7	0.0						
drop.Ib	19.4	8.4	0.0					
double.Ib	19.4	9.1	8.4	0.0				
situ.Ia	7.3	19.8	20.4	20.2	0.0			
lower.Ia	18.7	8.5	9.5	9.5	19.6	0.0		
drop.Ia	20.1	8.1	6.3	7.1	21.1	8.1	0.0	
double.Ia	19.1	9.5	8.9	8.7	19.9	9.3	8.0	...
...	...	...	...	...	...	...	...	...

Table 4.16: Upper lefthand corner of the distance matrix – step one of CA

Each of 17 the phenomena is thus compared to all other phenomena. Each comparison thus represents a dimension on which a given phenomenon is positioned in relation to the other phenomena. All these dimensions are combined, placing each phenomenon within 17-dimensional space. In order to visualise and interpret the data, the number of dimensions of this 17-dimensional space has to be reduced. This is the second step of CA: the high-dimensional space is reduced. Each dimension explains a portion of the variance in the data set, and all dimensions together account for 100% of the variation. We can examine the amount of variation that is represented by each dimension in a scree plot, in which for each dimension the amount of variance that is explained by that dimension, is given. The scree plot of the CA in this study is given in Figure

4.8.

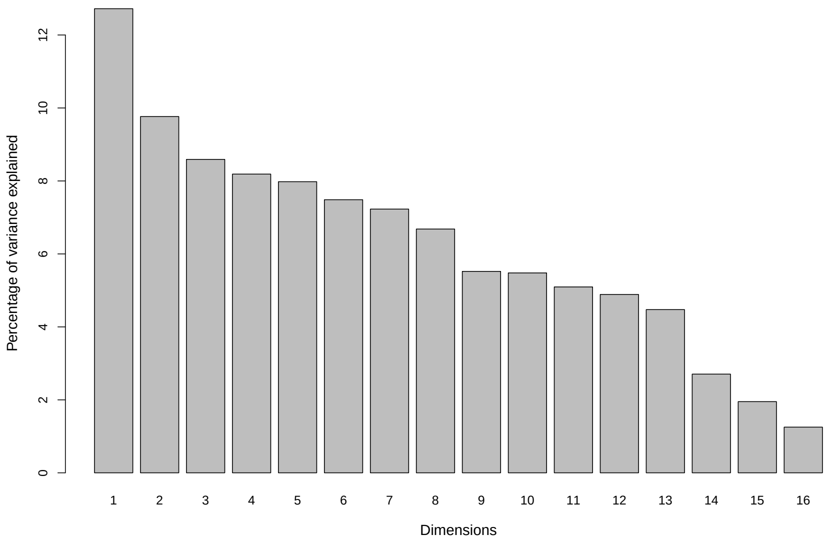


Figure 4.8: Percentage of variance explained per dimension of CA (case study I)

As can be seen in the scree plot, the first dimension explains the largest percentage of variance compared to the other dimensions. In the interpretation of the data, we want to keep the number of dimensions as low as possible (for easy interpretation and visualization), while at the same time keeping the percentage of explained variance as high as possible. In this respect, it is common practice to investigate only those dimensions that together explain the biggest portion of the variation. Deciding where to draw the line is not very easy when there is no clear point in the scree-plot after which every extra dimension adds only a very small percentage of explained variance. A decision has to be made, however, so let us look at the exact cumulative percentages of the first dimensions given in Table 4.17. As can be seen in the table, the first clear drop in explained variance is from dimension 1 to dimension 2, and the second from dimension 2 to dimension 3. As the cumulative percentage of explained variance of the three first dimensions is only 31.1%, however, it is preferable to look for a second drop in explained variance and set the cut-off point there. Let us draw the line after dimension 5: after this dimension, the percentage of explained variance is lower than 8%. By examining the first five dimensions, we get a picture of the dimensions that together represents 47.2% of the variance in the data set. Let us start by examining the first two dimensions. A two-dimensional plot of these dimensions is given in Figure 4.9.

Dimension	% of variance	Cumulative % of variance
Dimension 1	12.7	12.7
Dimension 2	9.9	22.6
Dimension 3	8.5	31.1
Dimension 4	8.1	39.2
Dimension 5	8.0	47.2
Dimension 6	7.5	54.7
Dimension 7	7.1	61.6
...	...	...
Dimension 16	1.3	100

Table 4.17: Percentage of explained variance per dimension (case study I)

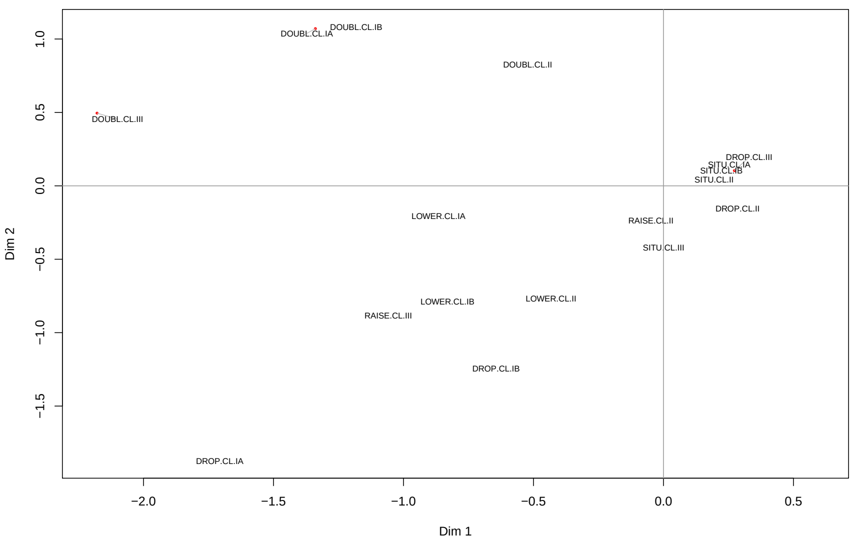


Figure 4.9: Dimensions 1 and 2 of CA (case study I)

Phenomena that are close to each other on the plot, are phenomena that have a similar distribution across the speakers (i.e. are more or less accepted by the same speakers), whereas phenomena that are far away from each other in the plot, are phenomena that have a dissimilar distribution across the speakers (i.e. there is little overlap in speakers who accept these phenomena). Note that phenomena that are located close to the origin (the point with coordinates (0,0) on the plot), are phenomena that do not contribute a lot to the variance in this dimension. In this two-dimensional plot, the x-axis represents the first dimension, while the y-axis represents the second dimension.

To interpret the first dimension, we have to look which phenomena are to the left of the y-axis, and which ones are to the right. In the plot we see that the following phenomena are to the right of the y-axis: *te*-in-situ in cluster types Ia, Ib, and II, and *te*-drop in cluster types II and III. On the y-axis itself we also find *te*-in-situ in cluster type III. The first dimension thus sets these phenomena apart from all other phenomena: *te*-drop in cluster types Ia and Ib, *te*-raising in all cluster types, *te*-lowering in all cluster types, and *te*-doubling in all cluster types. To interpret the second dimension, we have to determine which phenomena are above the x-ais, and which ones are below. In the plot we see that all *te*-doubling phenomena are located above the x-axis. Below the x-axis, all *te*-lowering groups together, together with *te*-raising in cluster type III, and even lower than that, we find *te*-drop in cluster types Ia and Ib. The second dimension thus sets *te*-doubling in all cluster types apart from *te*-lowering in all cluster types, *te*-raising in cluster type III and *te*-drop in cluster types Ia and Ib. Note that even though *te*-raising in cluster type II and *te*-in-situ in cluster type III are also slightly below the x-axis, they are very close to the origin, and therefore have not contributed a lot to the calculation of these two dimensions. Based on the first two dimensions, we can thus see that the phenomena cluster into three groups. The first group consists of the 'correct' versions of all clusters, *te*-raising in cluster type II, and *te*-drop in cluster types II and III: the most frequent phenomena which are accepted by more or less the same group of speakers. The second group of phenomena consists of *te*-doubling in all cluster types. The third group are all the *te*-lowering phenomena, together with *te*-raising in cluster type III, and *te*-drop in cluster types Ia and Ib. Note, however, that *te*-drop in cluster type Ib and specifically in Ia are located even lower than the *te*-lowering phenomena and *te*-raising in cluster type III. Let us therefore investigate the next two dimensions to see whether *te*-drop in cluster types Ia and Ib are set apart from the others in this group. The two-dimensional plot of the third and fourth dimension is given in Figure 4.10. Again, we are most interested in the position of the phenomena that are not too close to the origin. As was expected based on the findings of the first and second dimension, we see that the third dimension sets apart *te*-drop in cluster types Ia and Ib, located to the right of the y-axis, from the other phenomena with which it formed a group in the previous two dimensions: all *te*-lowering phenomena and *te*-raising in cluster type III are located to the left of the y-axis. The fourth dimension sets apart all phenomena in cluster type Ia from those in cluster type Ib, showing that if we dig deeper into the variation, these two clusters do not behave completely the same. This is not unexpected, as they do not consist of the same lexical elements, even though the selected position of *te* is the same. Furthermore, this dimension sets apart lowering in cluster types II and Ia. These phenomena are quite similar, but apparently at the fourth dimension they are separated.

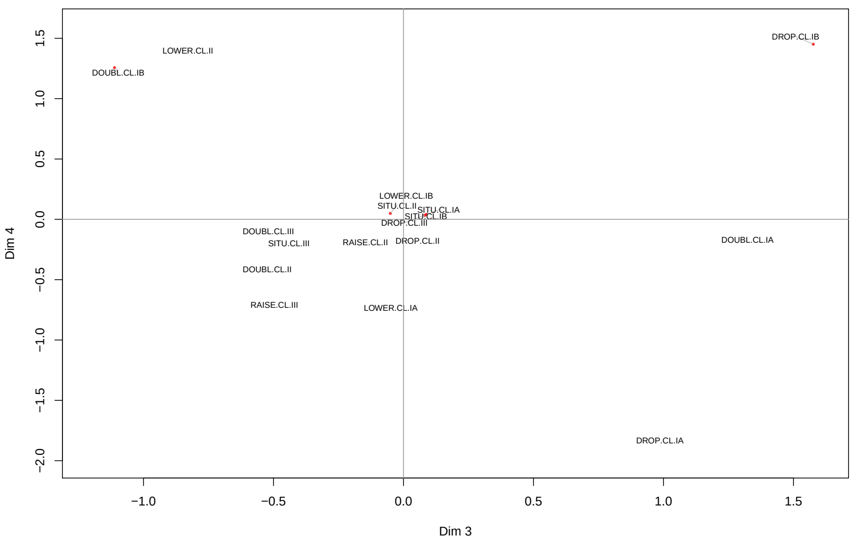


Figure 4.10: Dimensions 3 and 4 of CA (case study I)

So far we have seen that the first three dimensions together create four groups of phenomena: (i) *te*-doubling, (ii) *te*-lowering, and *te*-raising in cluster type III, (iii) *te*-drop in cluster types Ia and Ib, and (iv) all *te*-in-situ and *te*-drop in cluster types II and III. As it is slightly unexpected that *te*-lowering clusters together with *te*-raising in cluster type III, we might wonder whether these two are separated in the fifth dimension. Recall that we had decided to consider only the first five dimensions. We thus investigate the fifth dimension by plotting the fourth and fifth dimensions onto a two-dimensional plot, which is given in Figure 4.11.

Here, we only consider the y-axis, as we already interpreted dimension 4 (which is on the x-axis) in the previous plot. To interpret dimension 5, we thus have to look at what phenomena are above the x-axis, and which ones are below. We indeed see that *te*-raising in cluster type III is now below the baseline, while the *te*-lowering phenomena are above the baseline. The other observations we can make here are ones that we have already seen in the other dimensions, namely that *te*-drop in cluster type Ia behaves differently from *te*-drop in cluster type Ib and that *te*-doubling behaves differently from *te*-lowering.

Summing up, by examining the first three dimensions of the CA, we have found the following groups of phenomena: (i) *te*-doubling, (ii) all *te*-lowering,

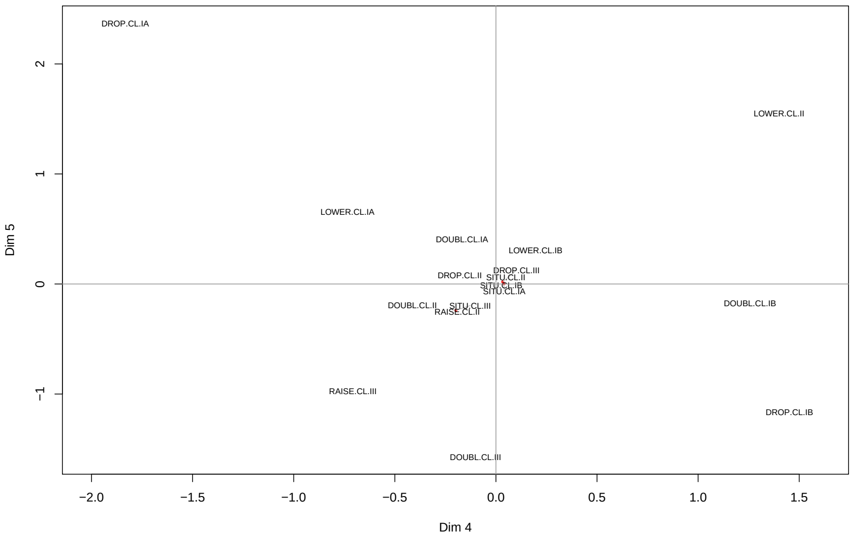


Figure 4.11: Dimensions 4 and 5 of CA (case study I)

and *te*-raising in cluster type III, (iii) *te*-drop in cluster types Ia and Ib, and (iv) *te*-in-situ, *te*-drop in cluster types II and III and *te*-raising in cluster type II. The fourth dimension separates *te*-lowering in cluster type II from *te*-lowering in cluster type Ia. The fifth dimension does the same for *te*-raising in cluster type III and *te*-lowering. *Te*-raising in cluster type III thus seems to be accepted by a group of speakers that until the fifth dimension is not set apart from the speakers who allow *te*-lowering. Let us now investigate whether we can get more insight into these findings by clustering the phenomena by means of a hierarchical clustering technique.

4.3.6.2 Hierarchical clustering: phenomena

A way to investigate which phenomena cluster together, is to use Hierarchical Clustering (henceforth HC) on the output of the CA.[11] That is, HC takes the output of CA, and groups phenomena together based on how similar they are. This is first represented in a hierarchical tree, as illustrated for the phenomena of this study in Figure 4.12. Objects are grouped together, and are thus similar, if there are few hierarchical steps to go from one object to the other in the tree (Husson et al. 2011). In order words, the phenomena are grouped together

[11]The calculations of this subsection were carried out in R (R Core Team 2014) with the HCPC function of the FactoMineR package (Husson et al. 2014).

in clusters based on the hierarchical tree, which can then be presented on a two-dimensional plot. However, one has to decide how many clusters to make, in other words, where to draw the line to cut off the groups made based on the hierarchical tree. As can be seen from Figure 4.12, the algorithm proposes to cut off the tree at the level of eight clusters: the big black horizontal line indicates the proposed cut-off point. However, dividing 17 phenomena over eight clusters might not be that informative, as quite a few clusters would contain only one phenomenon. In order words, such a clustering of phenomena is too fine-grained. This can be seen in the two-dimensional plot in which the phenomena are divided into eight clusters, as given in Figure 4.13.

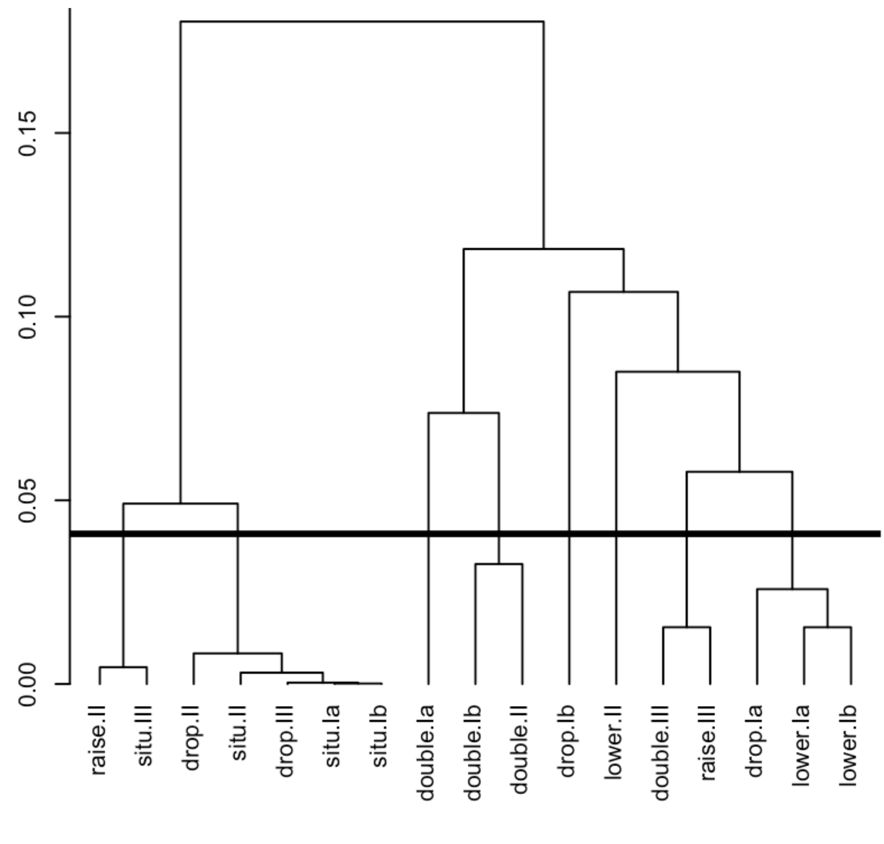


Figure 4.12: Hierarchical tree based on CA output (case study I)

As can be seen in the plot, five of these eight clusters contain only one phenomenon: the black cluster (*te*-drop in cluster type Ia), the green one (*te*-doubling in cluster type Ia), the pink one (*te*-drop in cluster type Ib), the blue one (*te*-doubling in cluster type Ib) and the grey one (*te*-lowering in cluster

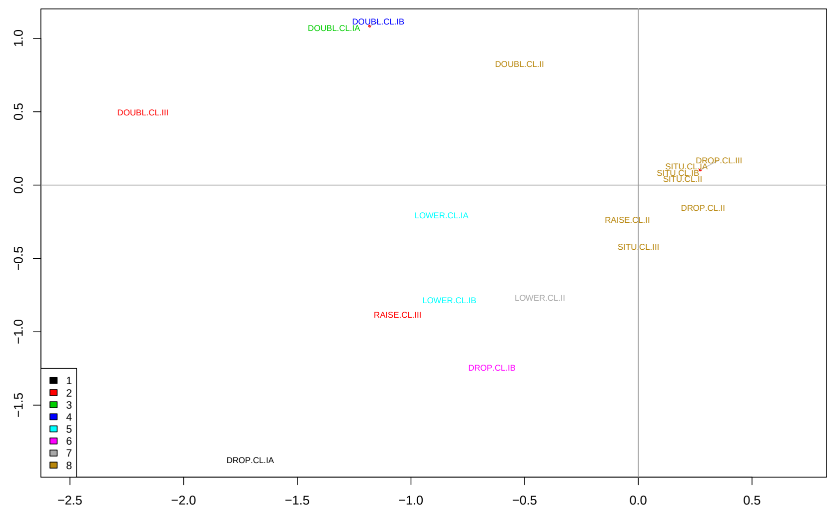


Figure 4.13: Phenomena clustered into eight groups (case study I)

type II). Let us therefore see whether clustering into seven or six groups yields more informative groupings of phenomena. The two-dimensional plot of HC with seven clusters is given in Figure 4.14, and with six clusters in Figure 4.15. As can be seen in Figure 4.14, if we divide the phenomena over seven clusters, there are still five clusters that contain only one phenomenon: the black cluster (*te*-drop in cluster type Ia), the pink one (*te*-drop in cluster type Ib), the green one (*te*-doubling in cluster type Ia), the blue one (*te*-doubling in cluster type Ib), and the grey one (*te*-lowering in cluster type II). This clustering is therefore still not that informative; it is still too fine-grained. Figure 4.15 shows that clustering into six clusters yields a much better result. Only three clusters contain just one phenomenon: the black cluster (*te*-drop in cluster type Ia), the blue one (*te*-drop in Ib), and the light blue one (*te*-lowering in cluster type II). All *te*-doubling phenomena are now put together (the red cluster), two *te*-lowering phenomena (in cluster types Ia and Ib) are grouped together with *te*-raising in cluster type III (the green cluster), and finally all *te*-in-situ and *te*-drop in cluster types II and III are grouped together (the pink cluster). This gives us the same picture as we arrived at by looking at the first five dimensions of CA in the previous subsection. Note that in all three clusterings we have seen so far, *te*-drop in cluster types Ia and Ib, and *te*-lowering in cluster type II are always in their own cluster. This means that regardless of the granularity of our clustering, these three phenomena are set apart from

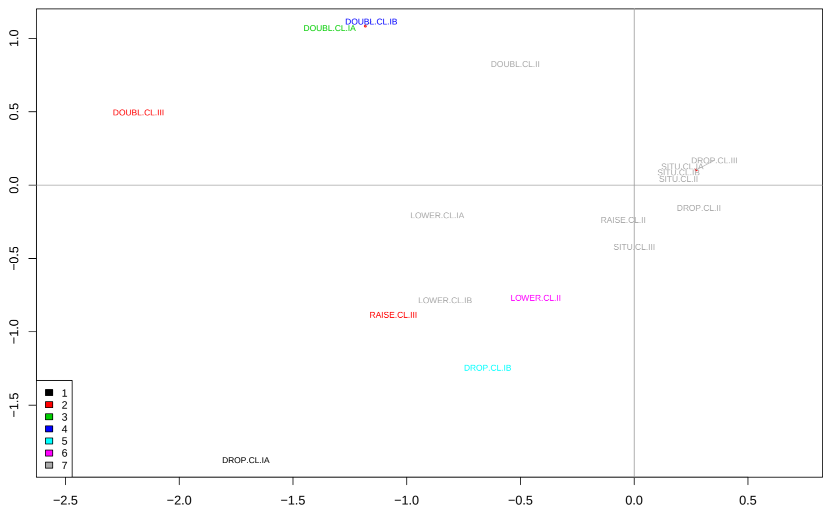


Figure 4.14: Phenomena clustered into seven groups (case study I)

the other phenomena. The picture remains the same if we restrict the HC to five clusters, the plot of which is given in Figure 4.16. When analysing this in formal terms in the next section, we thus have to keep in mind that these three phenomena are quite different from all the others. This is even more strongly the case for *te*-drop in cluster types Ia and Ib than for *te*-lowering in cluster type II. In particular, when we cluster into four or three clusters, *te*-lowering in cluster type II is grouped together with the other *te*-lowering phenomena, while *te*-drop in cluster type Ib remains a singleton cluster when the phenomena are divided into four clusters (Figure 4.17), and *te*-drop in cluster type Ia forms a single cluster even when the phenomena are divided into three clusters (Figure 4.18). *Te*-drop in cluster types Ia and Ib are thus the most distinct from all other phenomena and also from each other, as they are never grouped together into one cluster. *Te*-lowering in cluster type II, on the other hand, is slightly different from the other lowering phenomena, but is still grouped together into the same cluster from the granularity level of four clusters onwards. Note also that starting from the rather fine-grained level of six clusters to the more coarse-grained level of three clusters, all *te*-doubling phenomena are always in the same cluster, and never clustered together with other phenomena. In the formal analysis, we thus have to keep in mind that *te*-doubling is a uniform phenomenon across cluster types, and most likely has a quite distinct underlying mechanism from that of the other phenomena.

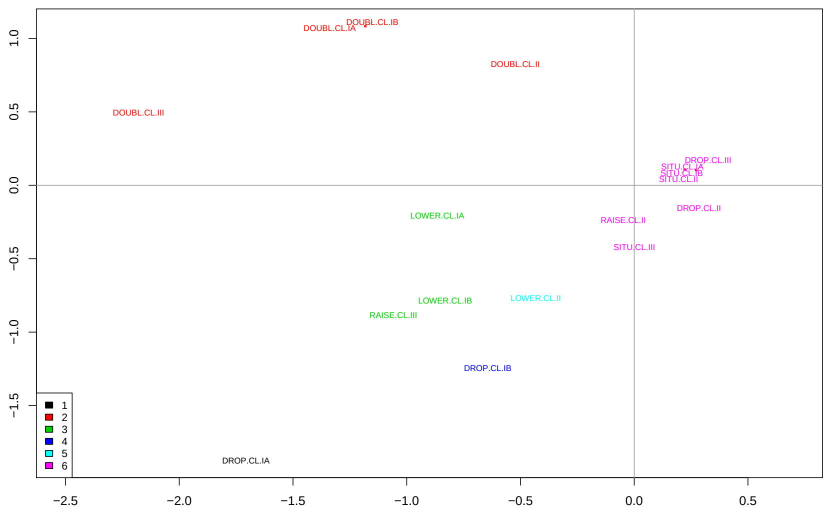


Figure 4.15: Phenomena clustered into six groups (case study I)

So far, we have seen that there are four quite robust groups of phenomena. One cluster consists all *te*-doubling. *Te*-drop in cluster types Ia and Ib are both singleton clusters. In subsection 1.3.4 we have seen however, that *te*-drop in cluster type Ia is only accepted by 8 speakers. Given this very low number of speakers, we might want to ignore this cluster, and focus on the more robust cluster of *te*-drop in cluster type Ib. The third cluster consists of all *te*-in-situ and of *te*-drop in cluster types II and III. The phenomena that are left are all *te*-lowering phenomena and *te*-raising in cluster type III. Perhaps unexpectedly, especially given that all *te*-doubling phenomena cluster together very consistently, we see that the *te*-lowering phenomena do not cluster together that robustly. That is, if we re-examine the two-dimensional plot in which the phenomena are divided into six clusters (Figure 4.15), we see that *te*-lowering in cluster types Ia and Ib is grouped together with *te*-raising in cluster type III, while *te*-lowering in cluster type II forms a singleton. It is only when we divide the phenomena into four clusters (Figure 4.17) that *te*-lowering in cluster type II is in the same cluster with the other lowering phenomena, and *te*-raising in cluster type III is in the same cluster as the other raising phenomenon (*te*-raising in cluster type II). It thus seems that of all the phenomena, *te*-lowering in cluster type II and *te*-raising in cluster type III behave somewhat differently from the same phenomena in the other clusters. This is again something that should be explained in the formal analysis in chapter 5. In addition, we can

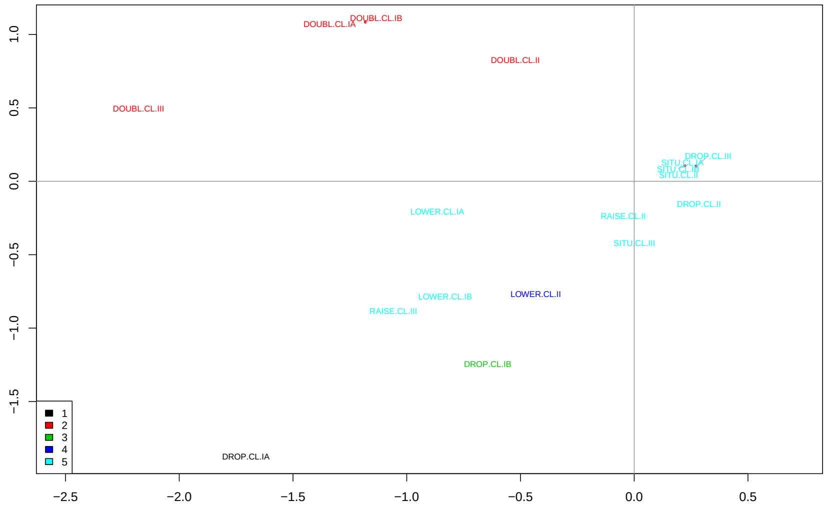


Figure 4.16: Phenomena clustered into five groups (case study I)

say that *te*-lowering in cluster types Ia and Ib form a robust cluster, which *te*-lowering in cluster type II only partly belongs to. *Te*-raising in cluster type III seems to behave in between *te*-lowering in cluster types Ia and Ib, and *te*-raising in cluster type II. Note that no matter how fine- or coarse-grained the clustering, we never end up with a cluster in which *te*-raising in cluster types II and III is grouped together without also including other phenomena. This is something we have to keep in mind in the formal analysis as well.

In summary, in this subsection we have seen that there are four robust clusters of phenomena:

1. *Te*-doubling in all cluster types
2. *Te*-drop in cluster type Ib
3. *Te*-in-situ in all cluster types, *te*-drop in cluster types II and III, and *te*-raising in cluster type II
4. *Te*-lowering in cluster types Ia and Ib

Now that we have more insight into how the phenomena cluster together, we can also see if there is any interesting geographical spread concerning these phenomena. This the topic of the next subsection.

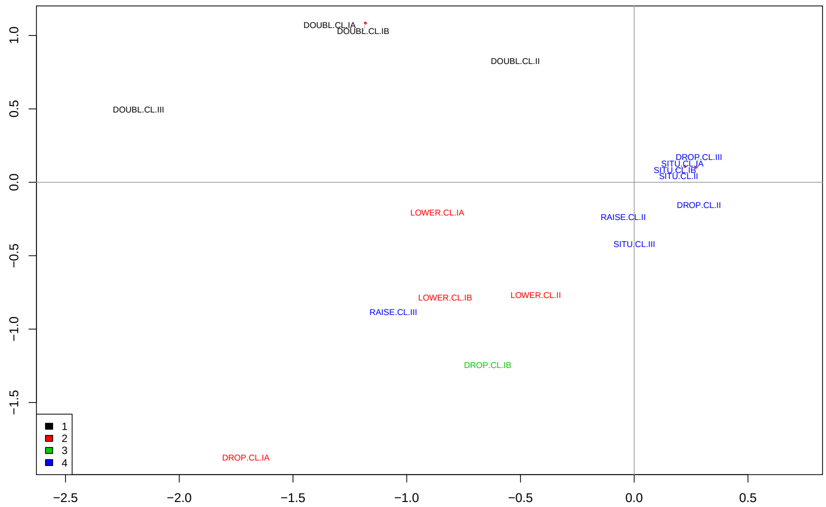


Figure 4.17: Phenomena clustered into four groups (case study I)

4.3.6.3 Hierarchical clustering: speakers

Recall that we have based the clustering algorithm on the output of CA. In 4.3.6.1, I showed how the first step of CA consists of calculating the distances between all phenomena in all cluster types (the rows of the raw data table), based on whether they are accepted by the same speakers or not (the columns of the raw data table). However, it is also possible to do the reverse. That is, we can also use CA based on the comparison of all speakers with one other (the columns of the raw data table), based on whether those speakers accept the same phenomena or not (the rows of the raw data table). If we use the output of this speaker-based CA as the basis for HC, the clustering algorithm will yield as output clusters of speakers rather than clusters of phenomena. The two-dimensional plot of the split into five clusters of speakers is given in 4.19.

This plot is not directly that informative, but what we can do is take a closer look at what defines these clusters by investigating their so-called ‘paragons’. Paragons are the individuals that are statistically the closest to the core of the cluster. In order words, they are the speakers that best represent the characteristics of the cluster they belong to (Husson et al. 2011). For each cluster type the phenomena that are part of the paragon’s grammar are given in Table 4.18.

Let us look at each cluster in turn. Cluster 1 is defined by two doubling phenomena, in cluster types Ia and II, it furthermore has the two ‘in-situ’

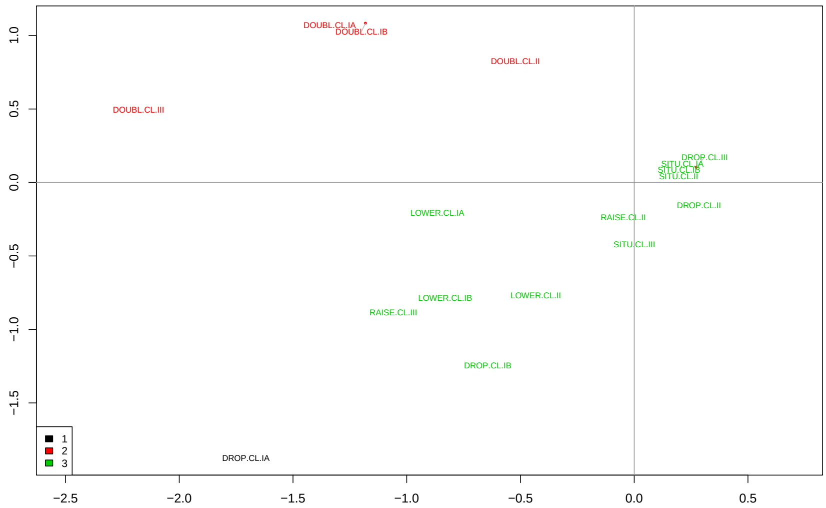


Figure 4.18: Phenomena clustered into three groups (case study I)

versions of these clusters, *te*-in-situ in cluster type Ib, and *te*-raising in cluster type II. If we look for the other two doubling phenomena in the entire Table, i.e. *te*-doubling in cluster types Ib and III, we see that they are not a characteristic of any of the five clusters, so we might consider cluster 1 to be the *te*-doubling cluster. We can furthermore observe that speakers who can double the *te* in a cluster also accept the 'in-situ' order. Lastly, it seems that speakers who double the *te* in cluster type II, also allow *te*-raising in this cluster.

Cluster 2 is defined by *te*-lowering in cluster types Ia and Ib. Let us consider this the '*te*-lowering in cluster type I'-cluster. As can be seen, speakers who lower in this cluster, also accept the 'in-situ' version. We furthermore see that it is characteristic of this cluster to also raise in cluster type II, but to allow *te*-in-situ in this cluster as well. Finally, *te*-in-situ and *te*-drop are both allowed in cluster type III.

Cluster 3 is defined by optionality regarding the position and presence of *te* in cluster type Ib: characteristic of this cluster is allowing *te*-drop, *te*-lowering and *te*-in-situ. The fact that *te*-in-situ in cluster types Ia and II is also listed here, is not that informative, as these are accepted by virtually all speakers in the data set. What is interesting, however, is that *te*-drop in cluster type II is characteristic for this cluster as well. We could hypothesise that *te*-drop in cluster type Ib is triggered by its presence in cluster type II, in which it is much more common in the entire language area in general. It might be an indication

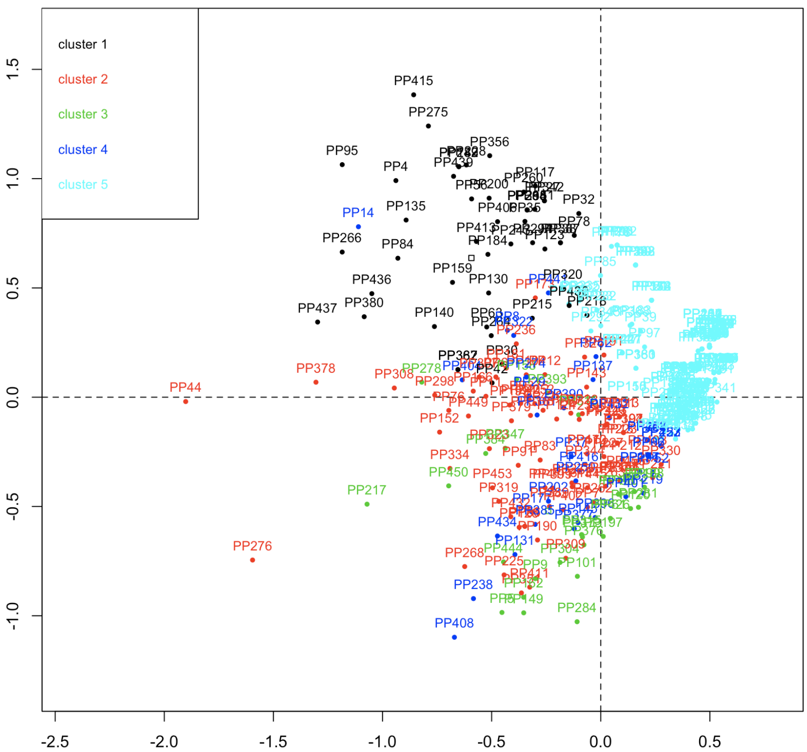


Figure 4.19: Speakers clustered into five groups (case study I)

that speakers who allow *te*-drop in cluster type Ib are overgeneralising the possibility of *te*-drop in cluster type II. However, as is discussed in 4.3.7, the weighted frequency of *te*-drop in cluster type Ib is in fact so low that we actually might consider it to be noise in the data.

Cluster 4 is defined by *te*-lowering in cluster types Ia and II. This cluster can thus be said to be a overal *te*-lowering cluster. Again, the *te*-in-situ phenomena that are listed are not very informative, as they are accepted by almost all or at least a significant subpart of the speakers. Cluster 5 is defined by all *te*-in-situ phenomena except for *te*-in-situ in cluster type III, but it does allow *te*-drop in this cluster. The last phenomenon that is characteristic of this cluster is *te*-raising in cluster type II. As *te*-drop with *zitten* (cluster type III) in three-verb clusters is generally considered to be more standard than *te*-in-situ (Haeseryn et al. 1997, Zwart 2007), we might want to consider this the most 'standard Dutch' cluster of all five. Note that *te*-raising in cluster type II is also characteristic for this cluster. We might thus consider this phenomenon to be characteristic of (a variety close to) the standard language as well. We can

Cluster	Paragon	Phenomena
Cluster 1	PP406	*te*-doubling in cluster type Ia *te*-in-situ in cluster type Ia *te*-doubling in cluster type II *te*-in-situ in cluster type II *te*-raising in cluster type II *te*-in-situ in cluster type Ib
Cluster 2	PP399	*te*-lowering in cluster type Ia *te*-in-situ in cluster type Ia *te*-lowering in cluster type Ib *te*-in-situ in cluster type Ib *te*-in-situ in cluster type II *te*-raising in cluster type II *te*-in-situ in cluster type III *te*-drop in cluster type III
Cluster 3	PP304	*te*-drop in cluster type Ib *te*-lowering in cluster type Ib *te*-in-situ in cluster type Ib *te*-in-situ in cluster type Ia *te*-in-situ in cluster type II *te*-drop in cluster type II *te*-raising in cluster type II
Cluster 4	PP250	*te*-in-situ in cluster type Ia *te*-lowering in cluster type Ia *te*-in-situ in cluster type Ib *te*-lowering in cluster type II *te*-in-situ in cluster type II *te*-in-situ in cluster type III *te*-drop in cluster type III
Cluster 5	PP1	*te*-in-situ in cluster type Ia *te*-in-situ in cluster type Ib *te*-in-situ in cluster type II *te*-raising in cluster type II *te*-drop in cluster type III

Table 4.18: Phenomena of representative speakers of the five clusters

summarize our description of the information in Table 4.18 as in Table 4.19.

It might be tempting to take cluster 2 and cluster 4 together as one bigger '*te*-lowering' cluster. As we have seen when discussing the clusters of phenomena created by HC in 4.3.6.2, however, *te*-lowering does not directly form a robust

Cluster	Description
Cluster 1	‘*Te*-doubling’ and *te*-raising in cluster type II
Cluster 2	‘*Te*-lowering in cluster type I’
Cluster 3	‘Opt. with *te* in cluster type Ib’ and *te*-drop in cluster type II
Cluster 4	‘*Te*-lowering’
Cluster 5	‘Standard language’ and *te*-raising in cluster type II

Table 4.19: Summary of characteristics of the five groups of speakers

cluster of phenomena. When it comes to *te*-lowering, both the CA and HC indicate that *te*-lowering in cluster type II is slightly different from *te*-lowering in cluster types Ia and Ib.

Finally, we can also map the five clusters onto the language area; that is, we can create a map and indicate for each speaker in each location to which of the five clusters this speaker belongs. The geographical distribution of the five clusters of speakers is given in Figure 4.20. As can be seen from the map, it is not the case that the five clusters are all in specific parts of the language area. From a dialectological point of view this is quite remarkable, as there are not many morphosyntactic phenomena in the Dutch language area that are not part of the standard language but that are so widespread. There are still some patterns to observe, however. First of all, it seems that the *te*-doubling cluster (the black cluster) is confined to the middle of the Netherlands, and occurs only sporadically in Belgium, Limburg and the northern part of the Netherlands. This is even more strongly the case for cluster 3 (the green cluster), which is characterised by optionality in the presence and position of *te* in cluster type Ib. As can be seen on the map, this type of speaker can only be found in the middle of the Netherlands and does not turn up in Belgium nor in the northern part of the Netherlands. Recall that the verb that selects *te* in this cluster is *beweren* ‘claim’. Apparently, for some speakers in the middle of the Netherlands the selectional requirements of this verb are looser than in the rest of the language area. The last observation that can be made is that the northern part of the Netherlands seems to be the most ‘well-behaved’ part with respect to the presence and position of *te*. That is, even though there are less observations for this part of the Netherlands in general, almost the entire area belongs to the ‘standard language’ cluster (the light blue cluster). With this, I conclude the discussion of the geographical spread of the phenomena.

4.3.6.4 Multiple Correspondence Analysis

In this last subsection, I use the exploratory statistical technique Multiple Correspondence Analysis (MCA), which is a specific application of the exploratory statistical technique CA (Husson et al. 2011:127). This is done to investigate the more fine-grained relationships between the phenomena, in which also the

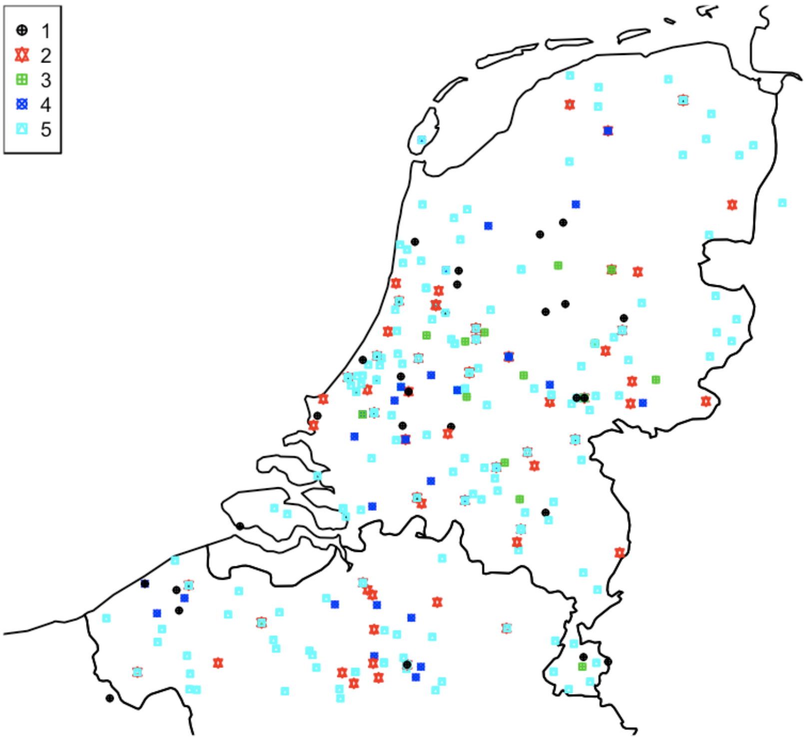


Figure 4.20: Geographical distribution of the five clusters of speakers (case study I)

different possibilities for *te*-raising (i.e. to V2 and to V1), *te*-lowering (i.e. to V2 and V3), and the three doubling configurations (i.e. *te*-V1-*te*-V2-V3, V1-*te*-V2-*te*-V3 and *te*-V1-V2-*te*-V3) are considered. The mechanisms behind MCA are comparable to those of CA. The main difference is that in CA needs binary encoding of the data, whereas MCA works with categorial data (Husson et al. 2011:127). This means that where for the execution of CA I have used a dataset in which all 1's, 2's and 3's were encoded as 0 and all 4's and 5' were encoded as 1, for the execution of MCA I use the 1 to 5 data encoding corresponding to the original ratings of the speakers. Furthermore, no test items are collapsed together into 'phenomena', as was done for the execution of CA above. This means that we have 28 test items (7 test items per cluster type, i.e. 7x4) that are compared with each other.[12] As in CA, in MCA the raw data table is inverted: the test items are rows, and the participants are columns. After the inversion of the raw data table, the same two steps are taken in the execution of MCA as I have explained above for CA in 4.3.6.1. First, the raw data table is converted into a distance matrix. To do so, the distance between every pair of test items is calculated. The more the same speakers give the same rating to the two test items, the smaller the distance is between these two test items. The distance matrix of the MCA is 28x28: al 28 test items are compared to all other test items. All these comparisons result in a 28-dimensional space. To visualise and interpret the data, the number of dimensions of this 28-dimensional space is reduced. Each dimension explains a part of the variance in the data set. All dimensions together explain 100% of the variation. Like I have done in 4.3.6.1, I look at a scree-plot in which for each dimension the percentage of explained variance is indicated to decide which dimensions to investigate. This scree-plot is given in Figure 4.21.

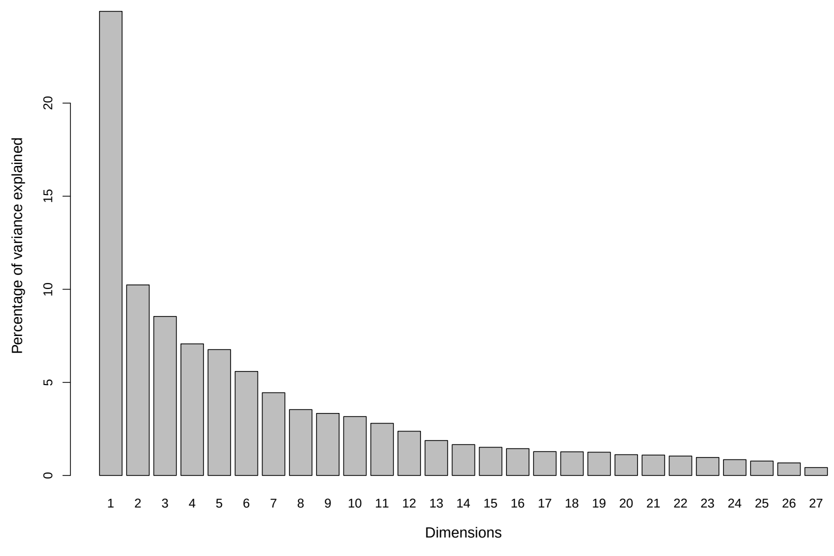


Figure 4.21: Percentage of variance explained per dimension of MCA (case study I)

[12]The 7 test items for cluster type Ia and Ib are: *te*-lowering to V2, *te*-lowering to V3, *te*-in-situ, *te*-drop, and all three doubling configurations. The 7 test items for cluster type II are: *te*-raising to V1, *te*-lowering to V3, *te*-in-situ, *te*-drop, and all three doubling configurations. The 7 test items for cluster type III are: *te*-raising to V2, *te*-raising to V1, *te*-in-situ, *te*-drop, and all three doubling configurations.

As can be seen from the scree-plot, the first dimension explains a very large portion of the variance in the data. I therefore decide only to investigate the first dimension. The plot presenting the first dimension (along with the second) is given in Figure 4.22.

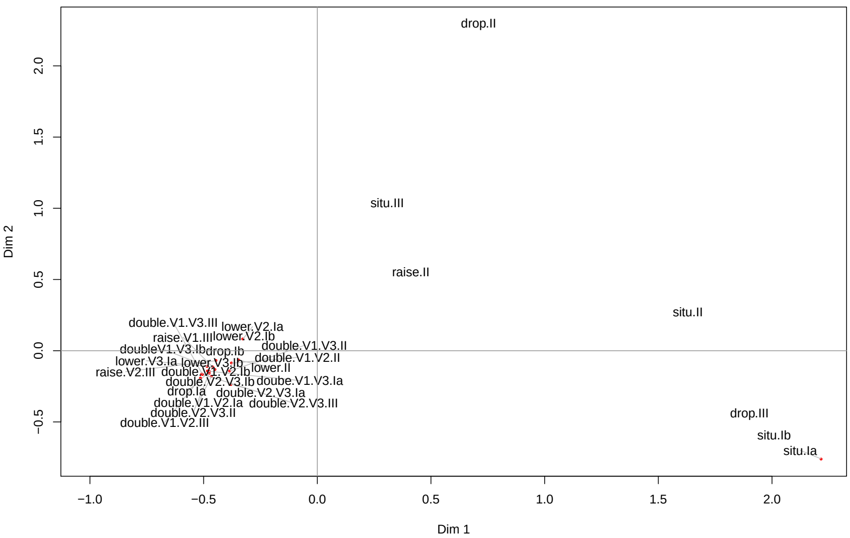


Figure 4.22: Dimensions 1 and 2 of MCA (case study I)

Recall from 4.3.6.1 that we have to interpret a plot like this as follows. The test items that are close to each other are the most similar to each other (i.e. received the same rating by more or less the same speakers). The test items that are far away from each other are the most dissimilar to each other (i.e. were rated differently by many of the speakers). Recall as well that test items that are located close to the origin (the point with coordinates (0,0) on the plot), are test items that do not contribute a lot to the variance in this dimension. In the two-dimensional plot in Figure 4.22, the x-axis represents the first dimension, while the y-axis represents the second dimension. Since we only investigate the first dimension, we only have to look at which test items are to the left of the y-axis, and which ones are to the right. As can be seen in the plot, a large group of test items are positioned close to the base: these test items thus did not contribute a lot to the variance of the first dimension. The test items that are positioned here are (i) all doubling configurations for all cluster types, (ii) both lowering to V2 and V3 for cluster types Ia and Ib, (iii) both raising to V2 and V1 in cluster type III, and (iv) the *te*-drop test items for cluster type Ia and Ib. Note that these are all test items which are grammatical for only a low percentage of speakers, and are furthermore always optional, see subsection 4.3.4. The first dimension sets apart this large group of test items against the test items with *te*-drop in cluster type II and III, *te*-raising in cluster type II, and *te*-in-situ in all cluster types. As we have seen in the previous subsections, *te*-raising in cluster type II is grouped together with all test items which can be considered as 'standard' Dutch. That is, it is grouped together with those test items in which *te* appears on the 'correct' position (*te*-in-situ), as well as with *te*-drop in cluster type II and III, which is mentioned as a 'standard' phenomenon with the *te*-selecting verbs in these clusters: *hoeven* 'need' and *zitten* 'sit' (see for example Haeseryn et al. (1997)).

4.3.7 Patterns to be explained in the formal analysis

In this subsection, I present the patterns as found in the data that should be explained in the formal analysis. First, it is necessary to draw a line somewhere between the frequencies I take to be indicative of the existence of a phenomena in a specific cluster type, and the ones that I regard as noise in the data. For this, I return to the weighted frequencies of all the phenomena per cluster type, as already presented in subsection 4.3.5, and repeated here as Table 4.20.

Cluster	*Te* in situ	*Te*-raising	*Te*-lowering	*Te*-drop	*Te*-doubling	Total
Type Ia	399 (86,9%)		30,7 (6,7%)	2,5 (0,6%)	26,8 (5,8%)	459 (100%)
Type Ib	348,3 (83,1%)		30,2 (7,2%)	18,1 (4,3%)	22,4 (5,4%)	419 (100%)
Type II	185,8 (46,8%)	76,8 (19,3%)	17,3 (4,4%)	88,9 (22,4%)	28,2 (7,1%)	397 (100%)
Type III	77,1 (19,6%)	30,4 (7,7%)		280,6 (71,0%)	6,9 (1,7%)	395 (100%)

Table 4.20: Weighted frequencies of all phenomena per type of cluster

Determining such a cut-off point will always be arbitrary to a certain extent, but in order to focus on the important and frequently occurring phenomena, I have decided to draw the line at a weighted frequency of 5%. A phenomenon in a specific cluster type which has a weighted frequency lower than 5% is considered as noise in the data in the formal analysis. This has the following consequences. *Te*-drop in cluster type Ia (0.6%) and Ib (4.3%) is now considered noise, as is *te*-lowering in cluster type II (4.4%), and *te*-doubling in cluster type III (1.7%).

With the noise separated from the rest of the patterns, we arrive at the following list of patterns that need to be explained by the formal analysis:

1. The highest degree of variation and optionality is found in cluster type II, with a lower degree in cluster type III, and the lowest degree in cluster type Ia and Ib
2. In cluster type Ia and Ib, *te* cannot be dropped
3. In cluster type Ia and Ib *te*-lowering and *te*-doubling occur very infrequently, and are always optional

4. In cluster type II, *te*-in-situ, *te*-raising and *te*-drop are all quite frequent phenomena, while *te*-doubling occurs infrequently in this cluster type

5. In cluster type III, the most frequent phenomenon is *te*-drop. Both *te*-in-situ and *te*-raising occur with lower to much lower frequencies, and *te*-doubling does not occur at all

6. There are three robust groups of phenomena:
   - *Te*-doubling in cluster type Ia, Ib and II
   - *Te*-lowering in cluster type Ia and Ib
   - *Te*-in-situ in all cluster types, *te*-drop in cluster type II and III, and *te*-raising in cluster type II

Many of the patterns as presented here have been replicated in a second questionnaire study with a different set of 500 native Dutch speakers. This second questionnaire investigated the presence and placement of *te* in the same three types of clusters, namely cluster type I in which the finite verb in Verb Second position selects the *te*-infinitive, cluster type II in which V1 *hoeven* 'need' in the non-finite three verb cluster selects the *te*-infinitive, and cluster type III in which V2 *zitten* 'sit' selects the *te*-infinitive.[13] The three test items are given in (15)–(17).

(15) Anne **zegt** haar oma gisteren lang [***te*** zijn$_1$ blijven$_2$ helpen$_3$].
Anne says her grandmother yesterday long to be remain help
'Anne says to have stayed to help her mom for a long time yesterday.'

(16) Johan zal morgen waarschijnlijk niet [**hoeven**$_1$ ***te*** gaan$_2$ werken$_3$].
Johan will tomorrow probably not need to go work
'Johan probably won't have to go to work tomorrow.'

(17) Eva zal wel weer de hele avond [hebben$_1$ **zitten**$_2$ ***te*** praten$_3$].
Eva will PRT again the entire evening have sit to talk
'Eva probably has been talking the entire evening again.'

The entire questionnaire can be found in Appendix B.[14] The same task and procedure were used as in questionnaire study I. The only difference was that

[13] Cluster type I was only tested with one test item, rather than with two (i.e. cluster type Ia and Ib of the first questionnaire study).

[14] Note that this questionnaire includes all six logically possible word orders of the non-finite three-verb clusters. For the comparison with the data of questionnaire study I, in which only the 1-2-3-order was tested, only the results of the 1-2-3-order are presented here. Discussing the frequencies of the phenomena in all cluster orders is beyond the scope of the thesis. I briefly return to the topic in chapter 6.

in the instruction text, the participants were informed about this being a very long questionnaire, and that would would therefore be divided into three parts.

The participants were recruited via the Meertens panel. This panel consists of a pool of native Dutch speakers which are subscribed to be invited for these type of studies. Their participation is voluntary. 563 native Dutch speakers completed the questionnaire, of which 500 were included for analysis. 19 participants were excluded based on having lived abroad for more than 10 per cent of their life. 44 participants were excluded based on inconsistent ratings of the filler items. The mean age of the participants was 58.5 years (*SD*=9.9, range=19-93). The gender division was 298 female, 202 male. 479 of the participants were born in the Netherlands, and 21 in Belgium.

The only difference concerning the between the design of questionnaire study I and II is that in the latter *te*-doubling was not tested. The weighted frequencies of *te*-in-situ, *te*-raising, *te*-drop and *te*-lowering for each cluster type are given in Table 4.21.

Type of cluster	*Te*-in-situ	*Te*-drop	*Te*-raising	*te*-lowering	Total
Cluster type I	226,3 (73,2%)	29,1 (9,4%)		54,2 (14,1%)	309 (100%)
Cluster type II	158,0 (45,9%)	86,6 (25,2%)	74,6 (21,7%)	24,8 (7,2%)	344 (100%)
Cluster type III	124,4 (31,0%)	223,5 (55,7%)	53,1 (13,3%)		401 (100%)

Table 4.21: Weighted frequencies of the phenomena – questionnaire study II

The data of questionnaire study II showed the following patterns. First, *te*-in-situ is the by far most frequent option for cluster type I, with low to very low frequencies for *te*-lowering and *te*-drop respectively. Second, in cluster type II, *te*-in-situ, *te*-drop and *te*-raising are all frequent phenomena, *te*-in-situ being the most frequent. *Te*-lowering in this cluster type is very infrequent. Third, in cluster type III, *te*-drop is the most frequent phenomena, with lower frequencies for *te*-in-situ and *te*-raising. Note that beside the weighted frequency of *te*-drop in cluster type I and *te*-lowering in cluster type II being above noise level (though still very infrequent) in this second questionnaire study, while they were below noise level in questionnaire study I, all other patterns are the same. The fact that most of main patterns of questionnaire study I have been replicated in questionnaire study II thus gives extra strength to the the general findings of this case study regarding the presence and placement of *te* across the three cluster types. Let us now move on to the analysis of the to be explained patterns, which is done in the next chapter.

CHAPTER 5

The analysis

5.1 Introduction

In this chapter, I present my formal analysis of the data that were presented as the data patterns to be explained in subsection 4.3.7. In section 5.2, I show that the *te*-selecting verbs in cluster type Ia and Ib as tested in the questionnaire study are lexical verbs, whereas *hoeven* 'need' and *zitten* 'sit' of cluster type II and III are semi-lexically used restructuring verbs. In section 5.3, I discuss previous analyses of the morphosyntactic status of *te*, and I present my own analysis. I propose that different types of *te*'s should be distinguished, one which is positioned on a functional head F and one which is the spell out of a specific feature on *v*. Based on these two proposals, I then present the syntactic structure of the test sentences of the questionnaire. In subsections 5.4.1-5.4.4, I present the analysis of the presence and placement of *te* in the different cluster types. In 5.4.5, I explain the groups of phenomena that were found in the statistical analyses of the previous chapter in light of my formal analysis. In 5.4.6, I present further evidence for the idea that the *te*-selecting verb of cluster type II, *hoeven* 'need' is in a transitional phase from the first stage of semi-lexicality to the second, whereas the *te*-selecting verb of cluster type III, *zitten* 'sit', unequivocally is at the first stage of semi-lexicality.

5.2 The *te*-selecting verbs in the cluster types

In this subsection, I show that the *te*-selecting verbs of cluster type II and III are semi-lexically used restructuring verbs, whereas those of cluster type Ia and Ib are lexical verbs. Let us start with the *te*-selecting verbs of cluster type II, *hoeven*, and cluster type III, *zitten*. Their respective clusters as they were tested in the questionnaire are repeated below in (1) and (2).

(1) Koen zal vanwege de winterstop vandaag niet [**hoeven**$_1$ *te*
Koen will because.of the winter.break today not need.INF to
gaan$_2$ voetballen$_3$].
go.INF play.football.INF
'Because of the winter break, Koen won't have to go and play football today.'

(2) Peter zal vanwege de nieuwe dienstregeling binnenkort nog langer
Peter will because.of the new train.schedule soon even longer
op de trein [moeten$_1$ **zitten**$_2$ *te* wachten$_3$].
on the train must.INF sit.INF to wait.INF
'Because of the new train schedule, Peter will soon have to wait even longer for the train.'

In chapter 2, I already presented Wurmbrand (2001)'s tests for the status of semi-lexically used restructuring verbs. Semi-lexically used verbs (i) occur in IPP form when embedded under a perfective auxiliary, (ii) do not allow extraposition, and (iii) they establish a thematic relation with the subject, and thus do not allow weather-*it* subjects. The examples in (3) and (4) show that both verbs occur in IPP form when embedded under a perfective auxiliary. In (5) and (6), we see that both verbs block extraposition of the embedded infinitive.

(3) ...dat Frans niet heeft { **hoeven** / * **gehoeven** } (te) werken.
...that Frans not has need.INF / need.PTCP to work
'...that Frans didn't need to work.'

(4) ...dat Frans niet heeft { **zitten** / * **gezeten** } (te) werken.
...that Frans not has sit.INF / sit.PTCP to work
'...that Frans hasn't been working.'

(5) *...dat Frans niet **hoeft** [de koek te eten].
...that Frans not need the cookie to eat
'...that Frans doesn't need to eat the cookie.'

(6) *...dat Frans **zit** [de koek te eten].
...that Frans sit the cookie to eat
'...that Frans doesn't need to eat the cookie.'

Semi-lexically uses verbs are predicted by Wurmbrand (2001) not to allow weather-*it* subjects. In (7), we see that *hoeven* does. As we can see in (8), *zitten* does not.

(7) ***Het*** **hoeft** niet te sneeuwen.
it need not to snow

'It doesn't need to snow.

(8) * ***Het*** **zit** niet te sneeuwen.
it sit not to snow
'It is not snowing.'

Based on these tests, it looks like *hoeven* is a functional restructuring verb rather than a semi-lexically used restructuring verb. There is one more property of *hoeven* that indicates that it behaves like other functional restructuring verbs in Dutch, such as the modal verbs, and that is the fact that it allows auxiliary switch. Auxiliary switch is a restructuring effect in which the auxiliary of the lower, lexical verb is selected instead of the auxiliary that is associated with the higher, functional verb (Cardinaletti and Shlonsky 2004). Let us first look at this phenomenon in Italian. It Italian, some verbs take the temporal auxiliary *avere* 'have', whereas others take *essere* 'be'. An example of the former is modal *volere* 'want', and of the latter *andare* 'go'. However, in cases of restructuring in which *volere* selects an infinitive like *andare*, the temporal auxiliary of the restructuring configuration can surface as *essere* 'be' rather than as *avere* 'have'. This is illustrated in (9).

(9) a. Ci ***avrei*** **voluto** andare con Maria.
there I.would.have wanted go.INF with Maria.
b. Ci ***sarei*** voluto **andare** con Maria.
there I.would.be wanted go.INF with Maria.
'I would have liked to go there with Maria.' (Cardinaletti and Shlonsky 2004:521-522)

In (9-a), *voluto* 'wanted' takes its own auxiliary: *avrei* 'would have'. As can be seen in (9-b), however, it is also grammatical for the auxiliary *essere* 'be' to be used, which is the auxiliary that normally combines with the lower lexical infinitive *andare* 'go'. In other words, in restructuring contexts, the modal or aspectual verb can be 'transparent' for auxiliary selection. Auxiliary switch is ungrammatical in non-restructuring contexts, as illustrated in (10).

(10) a. ***Avrei*** **detestato** andarci con Maria.
I.would.have hated go.INF with Maria.
b. * ***Sarei*** detestato **andar**ci con Maria.
I.would.be hated go.INF with Maria.
(Cardinaletti and Shlonsky 2004:521)

The verb *deterstare* 'hate' always selects temporal auxiliary *avere* 'have'. As can be seen in the example, only the use of this auxiliary, *avrei* 'have', is grammatical. A sentence in which the auxiliary associated with the lower lexical verb *andare* 'go' is used, namely *sarei* 'would be', is ungrammatical (10-b). While mostly restricted to Southern varieties, auxiliary switch is also attested in restructuring contexts in Dutch. Let us consider an example of Dutch auxiliary switch as well. Normally, Dutch *moeten* 'must' selects temporal auxiliary *hebben* 'have', and *gaan* 'go' selects temporal auxiliary *zijn* 'be'. In restructuring configurations, both the use of the auxiliary selected by modal *moeten*, namely *heb* 'have' (11-a), and the use of the auxiliary selected by the lower lexical infinitive *gaan*, namely *ben* 'am' are grammatical (11-b).

(11) a. …dat ik naar huis ***heb*** **moeten** gaan.
…that I to house have must.IPP go
'…that I had to go home.'
b. …dat ik naar huis ***ben*** moeten **gaan**.
…that I to house am must.IPP go
'…that I had to go home.'

Like in Italian, auxiliary switch is ungrammatical when the higher verb is a lexical verb. For example, the lexical verb *beloven* 'promise' takes the temporal auxiliary *hebben* 'have'. When *beloven* selects an embedded infinitive, like the infinitive *gaan*, only the use of the temporal auxiliary associated with *beloven*, namely *heb* 'have', is grammatical (12-a). The use of the temporal auxiliary associated with the embedded infinitive *gaan*, namely *ben* 'am', is ungrammatical (12-b).

(12) a. …dat ik ***heb*** **beloofd** naar huis te gaan.
…that I have promised.PTCP to home to go
'…that I promised to go home.'
b. *…dat ik ***ben*** beloofd naar huis te **gaan**.
…that I am promised.PTCP to home to go

The auxiliary switch test is not discussed by Wurmbrand (2001) to distinguish different types of restructuring verbs, because in German auxiliary switch is not attested. Given that Dutch modals allow auxiliary switch, however, it is interesting to test how *hoeven* and *zitten* behave with respect to this phenomenon. Recall from the examples discussed above that *hoeven* seemed to be more functional than *zitten*. We would thus expect that if one of the two verbs were to allow auxiliary switch, it would be *hoeven*. *Hoeven* indeed allows auxiliary switch, as is shown in (13-b). As for *zitten*, it is a little harder to test, as this verb generally does not select a motion verb, like *gaan* in (13), which is exactly the type of verb that is associated with *zijn* 'be' as auxiliary. This means that the non-auxiliary-switch variant with *zitten* in combination with *gaan* is also marginal (14-a). However, the auxiliary switch variant (14-b) is, in my opinion, even worse. Due to the marked nature of (14-a), however, we cannot determine conclusively whether the unacceptability of auxiliary switch with *zitten* is due to the auxiliary switch, or to the fact that this verb does not readily take a motion verb as its complement.

(13) a. …dat ik niet naar huis ***heb*** **hoeven** gaan.
…that I not to home have need.IPP go
'…that I didn't need to go home.'

b. . . . dat ik niet naar huis ***ben*** hoeven **gaan**.
. . . that I not to home am need.IPP go
'. . . that I didn't need to go home.'

(14) a. ??. . . dat ik naar huis ***heb*** **zitten** gaan.
. . . that I the book have sit.IPP go
'. . . that I have been reading the book.'
b. *. . . dat ik naar huis ***ben*** zitten **gaan**.
. . . that I to home am sit.IPP go

Summing up, the fact that *hoeven* allows auxiliary switch, and does not impose animacy or agentivity restrictions on the subject, makes it seem as if this verb has the same functional status as modals. On the other hand, modals never select a *te*-infinitive, and in that sense, *hoeven* is still semi-lexically used. It is clear, however, that *hoeven* is closer to being functional than *zitten*, which, by still imposing thematic restrictions on the subject, has the status of a semi-lexically used verb.

Having established that *hoeven* and *zitten* are semi-lexical restructuring verbs – though with different degrees of semi-lexicality – let us now look at the status of the *te*-selecting verbs in verb cluster types Ia and Ib, *zeggen* 'say' and *beweren* 'claim' respectively. Their verb clusters as they were tested in the questionnaire are repeated in (15) and (16).

(15) Anne **zegt** op haar comfortabele stoel [*te* willen$_1$ blijven$_2$ zitten$_3$].
Anne says on her comfortable chair to want.INF remain.INF sit.INF
'Anne says she wants to remain seated on her comfortable chair.'

(16) Stijn **beweert** met het geld van zijn erfenis die grote villa [*te* hebben$_1$ kunnen$_2$ kopen$_3$].
Stijn claims with the money of his inheritance that big villa to have can buy
'Stijn claims to have been able to buy that big villa with his inheritance money.'

According to Wurmbrand's (2001) classification for German, these verbs are non-restructuring verbs. This is also the case in Dutch. First, in (17) and (18), we see that these verbs cannot appear in IPP form. Second, the examples in (19) and (20) show that they allow extraposition (which is, in fact, the only option). Lastly, in (21) and (22), it is shown that they do not allow weather-*it* subjects.

(17) . . . dat ik heb { **gezegd** / * **zeggen** } te gaan.
. . . that I have say.PTCP / say.IPP to go
'. . . that I have said to go.'

(18) ... dat ik heb { **beweerd** / * **beweren** } te gaan.
... that I have say.PTCP / say.IPP to go
'... that I have said to go.'

(19) ... dat ik heb **gezegd** [een koekje te eten].
... that I have say.PTCP a cookie to eat
'... that I have said to eat a cookie.'

(20) ... dat ik heb **beweerd** [een koekje te eten].
... that I have say.PTCP a cookie to eat
'... that I have said to eat a cookie.'

(21) ****Het* zegt** niet te sneeuwen.
it says not to snow

(22) ****Het* beweert** niet te sneeuwen.
it claims not to snow

We can thus conclude that the *te*-selecting verbs of cluster types Ia and Ib are lexical (non-restructuring) verbs, whereas the *te*-selecting verb of cluster type II, *hoeven*, and cluster type III, *zitten* are semi-lexically used restructuring verbs. Furthermore, *hoeven* is close in morphosyntactic behaviour to a functional restructuring verb.

5.3 The position, function and morphosyntactic status of *te*

In this section, I start by discussing previous analyses of *te*. After that, I propose that we should distinguish between two sizes for the complement of *te*-selecting verbs. Then, I present my analysis of *te*, in which I argue that there are different types of *te*'s, which correspond to different syntactic sizes. After that, I give the underlying structures of the cluster types that were used as test sentences in the questionnaire study, and explain how the different *te*'s are spelled out.

5.3.1 Previous analyses of *te*

Three questions regarding the analysis of *te* are: (i) what position in the clause does it occupy?, (ii) what is its function?, and (iii) what is its morphological status? As will become clear in the brief discussion of previous analyses of *te*, no definitive answer has been provided yet for these three questions.

The most standard analysis of Dutch *te* is that it is an infinitival marker (i.e. its purpose is to mark the verb it appears on as an infinitive) which is positioned in T (a.o. Bennis and Hoekstra (1989), Den Besten and Broekhuis (1989), Beukema and Dikken (1989), Rutten (1991)). Furthermore, it is commonly assumed that *te* has the morphosyntactic status of a prefix. An argument from Bennis (2000) for the prefixal status of *te*, is that *te* and the infinitive cannot be separated (23-a), which mirrors the behavior of other verbal prefixes, like *ge-* (23-b).

(23) a. Ik beloof hem < * ***te*** > **op** < ***te*** > bellen.
I promise him to up to call
'I promise to call him.'
b. Ik heb hem < * ***ge-*** > **op-** < ***ge-*** > beld.
I have him GE- up- GE- called.
'I have called him.'

On the other hand, Zwart (1993) argues, based on the data in (24), that *te* cannot be a prefix, as it can scope over two infinitives (24-a). This is not possible with the prefix *ge-* (24-b) (Zwart 1993:104).

(24) a. Om in L.A. *te* **leven** en **sterven**.
for in L.A. to live.INF and die.INF
'To live and die in L.A.'
b. Om in L.A. ***ge-*** **boren** en *(***ge-***) **storven** te zijn.
for in L.A. GE- born and GE- died to be
'To be born and have died in L.A.'

Note that Bennis (2000) uses coordination constructions similar to the one in (24-a) as an argument in favor of the prefixal status of *te* (the judgment is that of Bennis 2000:115):

(25) De generaal moedigt het leger aan om ***te*** **strijden** en *(***te***) **winnen**.
the general encourages the army PRT for to fight and to win
'The general encourages the army to fight and win.'

IJbema (2001:70) agrees with the judgments of Zwart (1993) in (24-a), and argues based on that example that *te* is a clitic, given that clitics can have scope over two elements, whereas prefixes cannot (Miller 1991).[1] A second argument from Zwart (1993:103-104) against the prefixal status of *te* comes from the Groningen dialect, in which *te* can be separated from the infinitive by a bare noun (26) (Schuurman and Wierenga 1986:341).

(26) Dat hai begunt **te** ***kraant*** **lezen**.
that he begins to newspaper read
'That he starts to read the newspaper.' Groningen Dutch

However, these facts from Groningen Dutch do not hold for all Dutch varieties, which means that it could also be possible that in different varieties, *te* has a different status (as I have argued in Pots (2017)). At any rate, no conclusions

[1]Note, though, that the possibility of clitics having scope over two elements depends both on the type of clitic and the type of coordination, see Poletto (2000).

can be drawn regarding the morphosyntactic status of *te* for the entire language area based on an example from one specific dialect. So far, no conclusive evidence has been presented to make an empirically supported decision between the two proposals for the morphosyntactic status of *te*.

Let us return to the question about the function of *te*. Two objections against the standard analysis of *te* as an infinitival marker have come from Zwart (1993) and IJbema (2001). Zwart (1993) rejects this analysis because not all infinitives are marked by *te*; rather, it is the infinitival ending *-en* of Dutch infinitives that marks them as an infinitival verb. For example, modals and causatives never take a *te*-infinitive as complement. If *te* really were an infinitival marker, we would expect it to appear on all infinitival complements. (Zwart 1993:103) observes that *te* seems to mark a syntactic relation between the verb that selects the *te*-infinitive and the *te*-infinitive itself, rather than expressing tense. He therefore argues that *te* should be analysed as a complementiser or preposition. The former has also been proposed by Leys (1985:434). IJbema (2001) argues, however, that both the analysis of *te* as a complementiser and as preposition run into serious problems. While referring to her work for more elaborate argumentation, I will briefly discuss one argument against each analysis. One clear argument against *te* being a complementiser is that Dutch *om* already seems to fulfil the role of infinitival complementiser in complement clauses of verbs that select a *te*-infinitive. For example, the verb *proberen* can select either an *om te*-infinitive (27-a) or a *te*-infinitive (27-b).

(27) a. ...dat hij heeft **geprobeerd** ***om*** het boek ***te*** lezen.
...that he has tried in.order the book to read
b. ...dat hij heeft **geprobeerd** het boek ***te*** lezen.
...that he has tried the book to read
c. *...dat hij heeft **geprobeerd** ***om*** ***te*** het boek lezen.
...that he has tried in.order to the book read
'...that he has tried to read the book.'

The co-occurrence of two complementisers in one clause is an attested phenomenon in Dutch dialects (see van Craenenbroeck (2010)), so based on the presence of *om* we cannot say that *te* cannot be analysed as a complementiser. However, if one were to assume that *te* is a complementiser, we would not expect it to be separated from *om* by other material – given that they are both C-heads – which is clearly the case (27-a), and even obligatory so in (27-c). The analysis of *te* as a complementiser is thus not tenable.

As for the analysis of *te* as a preposition, one clear argument against that analysis is that *te* does not behave as preposition in its position in relation to separable particles (IJbema 2001:65-66). That is, prepositions always precede separable particles, as is illustrated in (28).

(28) Ik houd $<$ ***van*** $>$ **uit** $<$ * ***van*** $>$ gaan.
I love of out of go
'I love to go out.' (IJbema 2001:66)

Te, however, always follows the separable particle, as illustrated in (29).

(29) Ik besluit $<$ * ***te*** $>$ **uit** $<$ ***te*** $>$ gaan.
I decide to out to go
'I decide to go out.' (IJbema 2001:66)

Note, though, that *te* clearly can be used as a preposition, albeit with a slight archaic sound to it, as in (30).

(30) Ik ben geboren ***te*** **'s-Hertogenbosch**.
I am born to 's-Hertogenbosch
'I was born in 's-Hertogenbosch.'

As IJbema (2001) shows, the diachronic development of *te* started from a preposition. *Te* thus grammaticalised from a preposition – which is still in use in the language today – to an element that can appear on infinitives, depending on the verb that selects that infinitive.

Given that IJbema (2001) is the most recent analysis of *te*, I will briefly discuss her proposal, and its shortcomings. IJbema (2001) argues that *te* is indeed related to tense, but can also be a marker of Mood. She thus partly follows the standard analysis of *te* as an exponent of T, but proposes that *te* is an underspecified element that can be merged either in T or in Mood. Her main argument for *te* being positioned in T is that there are a number of verbs that can either select a bare infinitive or a *te*-infinitive, and depending on the presence or absence of *te*, the complement clause can or cannot be modified by a temporal adverb/adverbial phrase, or a modal verb. The verbs she discusses are *leren* 'teach', *helpen* 'help', *horen* 'hear', *voelen* 'feel' and *zien* 'see'. Note that for all these verbs, the semantics of either the complement or the verb itself changes when they select a *te*-infinitive compared to when they select a bare infinitive. For example, *leren + bare-infinitive* means 'teach somebody how to X' (31-a), whereas *leren + te-infinitive* means 'teach somebody that he should do X' (31-b) (Pardoen 1986:69). Similarly, *helpen + bare infinitive* means 'help somebody to do X' (32-a), whereas *helpen + te-infinitive* means 'help somebody in order for that somebody to be able to do X (in the future)' (32-b).

(31) a. Ik **leer** hem **lezen**.
I teach him read
'I teach him how to read.'

b. Ik **leer** hem ***te*** **lezen**.
I teach him to read
'I teach him that he should read.' (IJbema 2001:74)

(32) a. Ik **help** hem **lezen**.
I help him read
'I help him reading.'
b. Ik **help** hem ***te*** **lezen**.
I help him to read
'I help him so it will be possible for him to read.'

The other verbs, *voelen* 'feel', *horen* 'hear' and *zien* 'see', are verbs of perception when they select a bare infinitive, but have other meanings, namely 'think', 'ought to/be supposed to', and 'try/manage' respectively, when they select a *te*-infinitive, .[2] The contrastive pairs of *voelen*, *horen* and *zien* are given in (33)–(35) respectively.

(33) a. Ik **voel** de temperatuur **dalen**.
I feel the temperature drop
'I feel the temperature drop.'
b. Ik **voel** dit ***te*** **moeten** doen.
I feel this to must do
'I feel I should do this.'

(34) a. Ik **hoor** de klok **tikken**.
I hear the clock tick
'I hear the clock ticking.'
b. Ik **hoor** de hele dag ***te*** **werken**.
I hear the entire day to work
'I'm supposed to work the entire day.'

(35) a. Ik **zie** de trein **vertrekken**.
I see the train leave
'I see the train leave.'
b. **Zie** mij maar ***te*** **overtuigen**.
See me but to convince
'Try to convince me.'

IJbema (2001) shows for these verbs that only when they select a *te*-infinitive can the complement be modified by a temporal adverb/adverbial phrase, or embed a modal (cf. Aelbrecht (2009, 2010) concerning the temporal modification of the bare complements of Dutch modals). I illustrate both findings for *leren* 'teach'. In (36), we see that the *te*-less complement of *leren* cannot be modified

[2] It seems that the use of *zien* with the meaning 'try/manage' is not possible as present tense verb in declarative context (i).

(i) *Ik **zie** morgen **te** komen.
I see tomorrow to come
INTENDED: 'I'll try to come tomorrow.'

In (35-b), I have therefore given an example of an imperative. Given that a discussion of why this use of *zien* + *te-infinitive* is not possible in present tense declaratives is beyond the scope of this thesis, I leave this issue for future research.

by a temporal adverb (36-a), whereas this is fine when the complement of *leren* is a *te*-complement (36-b).

(36) a. * ***Vandaag*** **leer** ik hem [zich ***morgen*** verdedigen].
today teach I him REFL tomorrow defend
b. ***Vandaag*** **leer** ik hem [zich ***morgen*** te verdedigen].
today teach I him REFL tomorrow to defend
'Today I teach him to defend himself tomorrow.'

In (37), we see that the *te*-less complement of *leren* cannot contain a modal (37-a), whereas this is grammatical when the complement is a *te*-complement (37-b).

(37) a. *Ik **leer** hem [***moeten*** lezen].
I teach him must read
b. Ik **leer** hem [***te moeten*** lezen].
I teach him to must read
'I teach him that he should read.'

IJbema (2001) presents these data as evidence that *te* is merged in T. I have two arguments against her conclusion. First, the fact that *te*-infinitives can be modified by a temporal adverb/adverbial phrase and embedded a modal, whereas bare infinitives cannot, shows that *te*-infinitives must contain a T head, but this does not mean that *te* is merged in this head. It merely shows that (some, see below) *te*-complements have a tense layer. Second, IJbema (2001) discusses the difference between bare and *te*-infinitive complements of a specific set of verbs that can select both, but in which the meaning of the selecting verb or the complement changes depending on the type of complement. There are also verbs, however, that can only select a *te*-infinitive, but do not allow modification by adverbials/modals of that infinitive, like *beginnen* 'begin', *durven* dare or *zitten* 'sit'. This is illustrated in the following examples.

(38) a. * ***Vandaag*** **begin** ik [***morgen te*** lezen].
today begin I tomorrow to read
b. *Ik **begin** [***te*** **moeten** lezen.]
I begin to must read

(39) a. * ***Vandaag*** **durf** ik [***morgen te*** lezen].
today dare I tomorrow to read
b. *Ik **durf** [***te*** **moeten** lezen.]
I dare to must read

(40) a. * ***Vandaag*** **zit** ik [***morgen te*** lezen].
today sit I tomorrow to read
b. *Ik **zit** [***te*** **moeten** lezen.]
I begin to must read

These data thus suggest that not all *te*-complements have a T layer, and therefore that *te* is not – or at least not uniformly – merged in T. Moreover, the difference between the data in (36)–(37) on the one hand and those in (38)–(40) on the other suggests that we might be dealing with different sizes of *te*-complements in the two groups of verbs (see also Cremers (1983)). I return to this idea below, when presenting my analysis of *te*.

IJbema (2001) proposes that *te* can be merged either in T or in Mood. Let us briefly go over her argument for *te* sometimes being merged in Mood, which is based on data from West Flemish. She starts from the following examples, which are taken from (Haegeman 1995:53-59).

(41) ...da Jan willen$_2$ Valère nen boek geven$_3$ **een**$_1$.
...that John want.INF Valère a book give.INF has.FIN
'...that John has wanted to give Valère a book.'

(42) ...da Jan **oa**$_1$ willen$_2$ Valère nen boek geven$_3$.
...that John had.FIN want.INF Valère a book give.INF
'...that John had wanted to give Valère a book.'

In (41), the finite present tense auxiliary *een* 'has' occurs in cluster final position. By contrast, in (42), the past tense form *oa* 'had' occurs in cluster initial position. According to Haegeman (1995), *oa* in this position receives a modal, irrealis interpretation. The cluster final versus cluster initial position of present tense *een* versus past tense *oa* is a preference rather than an absolute judgment (Haegeman 1995). In non-finite contexts, the cluster initial position is unavailable for the perfective auxiliary.

(43) Mee Valère **te** willen$_2$ dienen boek kuopen$_3$ ***een***$_1$.
with Valère to want.INF that book buyINF have.INF
'Valère having wanted to buy that book.'

(44) *Mee Valère **te** ***een***$_1$ willen$_2$ dienen boek kuopen$_3$.
with Valère to have.INF want.INF that book buyINF

Haegeman (1995) argues that the perfective auxiliary in sentences like (42) has moved from its lower merge position to a higher functional projection F1. According to her, this is also the merge position of *te*, which accounts for the ungrammaticality of (44): *te* and *een* compete for the same functional head position. Recall that it is preferred for *oa* 'had' to be in cluster initial position (42). Given that *oa* in this position receives a modal, irrealis interpretation, and that this position is unavailable when *te* is present in the clause (44), both Haegeman (1995) and IJbema (2001) conclude that the merge position of *te* in (43) is a functional head associated with irrealis mood. As IJbema (2001) adopts Cinque (2001)'s functional sequence, she concludes that *te* can be merged in Mood$_{irrealis}$.

I would like to briefly discuss three arguments against IJbema (2001)'s conclusion that *te* can be merged in Mood$_{irrealis}$. First, this conclusion is based on the assumption that (44) is ungrammatical because *te* and *een* compete for the same head position, combined with the observation that *oa* 'had', which has a modal, irrealis interpretation, is moved to this position in finite contexts (42). However, there might be other reasons for the ungrammaticality of (44), which are not discussed in IJbema's analysis. It might be the case that infinitival *een* cannot move for other reasons, but IJbema (2001) does not discuss a West Flemish example of a non-finite verb cluster with *een* as V1 and without *te*. Such an example should be grammatical based on her analysis, as these elements compete for the same position, but without *te*, *een* should be able to move. Without such examples, we cannot definitively conclude that *te* and *een* compete for the same position. Moreover, in the grammatical non-finite clause in (43), *een* should crucially not receive a modal, irrealis interpretation, given that it has not moved to Mood$_{irrealis}$. Neither Haegeman (1995) nor IJbema (2001) discuss whether this is the case. Second, even if we would accept the analysis of *te* as being merged in a position associated with irrealis Mood in West Flemish, there is no reason to assume that this holds for all Dutch varieties. For example, the Standard Dutch version of the grammatical West-Flemish non-finite cluster in (43) has a very different word order.

(45) Met Valère dat boek **te *hebben***$_1$ willen$_2$ kopen$_3$.
with Valère that book to have.INF want.INF buy.INF
'Valère having wanted to buy that book.'

In (45), it is clear that *te* and *hebben* 'have' do not compete for the same position. Third, IJbema (2001) analyses *te* as a clitic. Clitics normally do not block verb movement, so if *te* is a clitic, it is very strange that it would block movement of an auxiliary, as in (44). Summing up, even if we accept the analysis of *te* as being merged in Mood$_{irrealis}$ in West-Flemish based on the examples discussed by IJbema (2001), we cannot generalise this analysis to the entire language area.

A last part of IJbema (2001)'s analysis I want to focus on, is her proposal that there are two sizes of *te*-infinitives. Cremers (1983) was the first to propose a split among Dutch infinitives. He argues that there is a group of *te*-infinitives that are timeless, i.e. that do not independently refer to time, and a group of *te*-infinitives that do have their own temporal reference. He proposes that the former have the size of a VP, whereas the latter have the size of a CP (labeled 'S' in this work). Examples of verbs that fall into the first group are given in (46), and examples of verbs that fall into the second group in (47).

(46) Verbs that take a VP complement (Cremers 1983)
proberen 'try', *durven* 'dare', *kunnen* 'can', *moeten* 'must', *dwingen* 'force'.

(47) Verbs that take a CP complement (Cremers 1983)
zeggen 'say', *denken* 'think', *beweren* 'claim', *beseffen* 'realise', *meedelen* 'announce'

Note that the first group of verbs contains both verbs that select *te*-complements (*proberen*, *durven*, *dwingen*) and verbs that select bare complements (*kunnen*, *moeten*). According to Cremers (1983), the complement of the first group of verbs cannot independently refer to time, whereas this is possible in the complement of the second group.[3] Furthermore, the complement of the first group of verbs cannot contain the modal *zullen* 'will', whereas the complement of the second group can. Consider the following examples.

(48) a. * ***Vandaag*** **probeert** Jan je ***morgen*** te bellen.
today tries Jan you tomorrow to call
b. ***Vandaag*** **zegt** Jan jou ***morgen*** te bellen.
today says Jan you tomorrow to call
'Today Jan tells you he will call you tomorrow.' (Cremers 1983:182-3)

(49) a. *Jacoba **probeert** jou te ***zullen*** bezoeken.
Jacoba tries you to will visit
b. Jacoba **zegt** jou te ***zullen*** bezoeken.
Jacoba says you to will visit
'Jacoba says that she will visit you.' (Cremers 1983:181)

The complement of *proberen* 'try' cannot contain a modal referring to the future (49-a), which is possible in the complement of *zeggen* (49-b). *Proberen* and its complement infinitive cannot be independently modified by conflicting temporal adverbs, while this is possible in the case of *zeggen*. Based on these differences, Cremers (1983) concludes that the complement of verbs like *proberen* from a temporal unit with that verb, whereas the complement of verbs like *zeggen* do not. However, IJbema (2001) rejects Cremers's proposal that the complement of the first group of verbs is a VP, and the complements of the second group of verbs a CP. Instead, she argues that the verbs that take a *te*-infinitive should be split up into two groups (thus leaving aside verbs that select a bare infinitive): the first group selects a complement that projects only up to Mood$_{irrealis}$P and which is irrealis, and the second a complement that projects up to TP and which is realis.[4] Examples are given in (50) and (51) respectively.

(50) Verbs that select an irrealis (Mood$_{irrealis}$P) complement (IJbema 2001)
adviseren 'advise', *beloven* 'promise', *besluiten* 'decide', *denken* 'intend', *durven* 'dare', *proberen* 'try', *weigeren* 'refuse'

(51) Verb that select a realis (TP) complement (IJbema 2001)
beseffen 'realise', *beweren* 'claim', *meedelen* 'announce', *zeggen* 'say'

In her analysis *te* is positioned in Mood$_{irrealis}$ in the complement of the first group of *te*-selecting verbs, and in T in the complement of the second group.

[3] See Aelbrecht (2009, 2010) for counterexamples to this claim.
[4] In Cinque (1999)'s hierarchy, Mood$_{irrealis}$ is positioned higher than T.

IJbema (2001) presents the following counterargument against Cremers's claim that the complement of a verb like *proberen* is a VP (see (46)). The argument is based on the following example from Pardoen (1986).

(52) Ik heb ***gisteren*** **geprobeerd** hem ***vandaag*** niet te hoeven ontmoeten.
I have yesterday tried him today not to need meet
'Yesterday I tried not having to meet him today.' (Pardoen 1986:54)

Given that in this example *proberen* is independently modified by a temporal adverb that contrasts with the temporal adverb that modifies the complement verb, IJbema (2001) concludes that Cremers (1983)'s analysis of verbs like *proberen* as taking a VP complement is untenable. I take issue with this conclusion. As already discussed by (Wurmbrand 2001:312-315) for its German cognate *versuchen*, the complement of the verb *proberen* can be of different sizes. For Dutch, this is evidenced by the position of the embedded object, the morphological form of *proberen* itself (past participle versus IPP form), and the acceptability of *om* 'for' in its complement.

(53) a. …dat ik heb ***geprobeerd*** (***om***) ***hem*** te bellen.
…that I have tried.PTCP for him to call
b. …dat ik ***hem*** heb ***geprobeerd*** (* ***om***) te bellen.
…that I him have tried.PTCP for to call
c. …dat ik ***hem*** heb ***proberen*** (* ***om***) te bellen.
…that I him have tried.IPP for to call
'…that I have tried to call him.'

As can be seen from the contrast between (53-a) and (53-b), the presence of *om* blocks object scrambling: the pronoun *him* can only appear to the left of *geprobeerd* when *om* is absent. Given that *om* is an infinitival complementiser, the size of the complement in (53-a) is a CP. The size of the complement in (53-b) must be smaller than a CP, and this is also the case in (53-c). The size of the complement in the latter example is probably even smaller, given that IPP only occurs with functional and semi-lexical restructuring verbs (cf. chapter 2, Wurmbrand (2001)). Returning to IJbema (2001)'s example in (52), we see that the embedded object is not scrambled here. This means that the size of the complement is identical to that of *geprobeerd* in (53-a): a CP. We thus expect that independent temporal modification of *geprobeerd* and its complement should be possible. If we try the same with a sentence in which the embedded object is scrambled or in which *proberen* occurs in its IPP form, the result is ungrammatical (54).

(54) a. *Ik heb ***gisteren*** hem **geprobeerd** ***vandaag*** niet te hoeven ontmoeten.
I have yesterday tried him today not to need meet

b. *Ik heb ***gisteren*** hem **proberen** ***vandaag*** niet te hoeven ontmoeten.
I have yesterday tried him today not to need meet
INTENDED: 'Yesterday I tried not having to meet him today.'

In other words, we should beware of the fact that *proberen* can take different sizes of complements, and that is therefore not the ideal verb to argue for or against a particular complement size if we only consider one of its possible complements. In the next subsection, I present my own proposal regarding the different types (and therefore sizes) of *te*-complements.

5.3.2 Different sizes of *te*-complements

Based on the preceding discussion, I want to stick with Cremers's (1983) proposal that there are a number of *te*-selecting verbs that behave similar to verbs that select bare infinitives, and that other *te*-selecting verbs behave differently, specifically in selecting a larger complement. This split is also in line with Wurmbrand (2001)'s restructuring versus non-restructuring infinitives (see also Chapter 2).[5] Some of the verbs that according to IJbema (2001) take a Mood$_{irrealis}$P complement (e.g. (50)), and all of the verbs that she argues take a TP complement (e.g. (51)) fall in the latter group. The *te*-selecting verbs that behave similarly to verbs that select bare infinitives are given in (55). The *te*-selecting verbs that take a larger complement (TP or CP) are illustrated in (56).

(55) Verbs that select a small complement
hoeven 'need', *zitten* 'sit', *staan* 'stand', *liggen* 'lie', *lopen* 'walk', *durven* 'dare', *beginnen* 'begin' and *proberen* 'try'

(56) Verb that select a large complement (TP or CP)
zeggen 'say', *beweren* 'claim', *besluiten* 'decide', *adviseren* 'advise', *beseffen* 'realise' et cetera

A way to distinguish the two groups is to identify whether the verbs can also

[5]Note that in the group of verbs that select a larger complement, a second subdivision can be made between verbs that select an irrealis complement on the one hand, and those that select a factive or propositional complement on the other. Both subtypes are non-restructuring verbs, but the former are 'reduced non-restructuring verbs' and the latter 'full non-restructuring verbs'. I refer the reader to Wurmbrand (2001) for discussion of this aspect of restructuring, and to Wurmbrand and Lohninger (2019) for details and discussion of the three classes of verbs (their Attitude, Irrealis and Tenseless classes) more generally, and from a cross-linguistic perspective.

take complement other than a *te*-infinitive. All the verbs in the second group can also take an alternative complement that is demonstrably a CP, namely either an *om te*-complement or a finite *dat*-complement, as illustrated in (57)–(61). This is not possible for the verbs given in (55), as shown in (62)–(66).[6,7]

(57) a. Ik **zeg** ***te*** zullen werken.
I say to will work
b. Ik **zeg** ***dat*** ik zal werken.
I say that I will work
'I'm saying that I'll work.'

(58) a. Ik **beweer** ***te*** zullen werken.
I claim to will work
b. Ik **beweer** ***dat*** ik zal werken.
I claim that I will work
'I'm claiming that I'll work.'

(59) a. Ik **besluit** ***te*** werken.
I decide to work
b. Ik **besluit** ***om*** te werken.
I decide for to work
'I decide to work.'

(60) a. Ik **adviseer** je ***te*** werken.
I advise you to work
b. Ik **adviseer** je ***om*** te werken.
I advise you for to work
'I advise you to work.'

(61) a. Ik **besef** ***te*** moeten werken.
I realise to must werken
b. Ik **besef** ***dat*** ik moet werken.
I realise that I must work
'I realise that I have to work.'

[6] For *hoeven*, I give both positions for negation in the ungrammatical examples, to make sure the ungrammaticality is not the result of the wrong positioning of the negation. For the posture verbs, *zitten*, *staan*, *liggen*, I only give an example for *zitten*; they all behave the same. I do not give an example for *proberen*, as I already discussed that it can also take a larger *om te*-complement in the examples in (53). See also Sharvit (2003) and Wurmbrand and Lohninger (2019) on the 'in-between' status of the verb *try* in English, and cross-linguistically.

[7] This observation also holds for English verbs that fall in the Tenseless class of Wurmbrand and Lohninger (2019); these verbs never take a finite complement (i-c), whereas the verbs of the Attitude (i-a) and Irrealis (i-b) class can.

(i) a. Clara *claimed* to have left/that he left. (Attitude)
b. Clara *decided* to leave/that he would leave. (Irrealis)
c. Clara *tried* to win/??that she would win. (Tenseless) (Wurmbrand and Lohninger 2019:18)

(62) a. Ik **hoef** niet ***te*** werken.
I need not to work
'I don't need to work.'
b. *Ik **hoef** < niet > ***om*** < niet > te werken.
I need not for not to work
c. *Ik **hoef** < niet > ***dat*** ik < niet > werken.
I need not that I not work

(63) a. Ik **zit** *te* werken.
I sit to work
'I'm working.'
b. *Ik **zit** ***om*** te werken.
I sit for to work
c. *Ik **zit** ***dat*** ik werk.
I sit that I work

(64) a. Ik **loop** ***te*** werken.
I walk to work
'I'm working.'
b. *Ik **loop** ***om*** te werken.
I walk for to work
c. *Ik **loop** ***dat*** ik werk.
I walk that I work

(65) a. Ik **durf** ***te*** werken.
I dare to work
'I dare to work.'
b. *Ik **durf** ***om*** te werken.
I dare for to work
c. *Ik **durf** ***dat*** ik werk.
I dare that I work

(66) a. Ik **begin** ***te*** werken.
I begin to work
'I start to work.'
b. *Ik **begin** ***om*** te werken.
I begin for to work
c. *Ik **begin** ***dat*** ik werk.
I begin that I work

To show that all verbs that select a large complement, can also take either an *om te*-complement or a finite *dat*-clause, I have adapted (IJbema 2001:151)'s Table 1 (of Appendix C), as in Table 5.1 below. This table gives an overview of the relevant properties of all *te*-selecting verbs.[8] There are also a number

[8] I have left out a number of archaic verbs that are part of IJbema (2001)'s Table 1, namely *pogen* 'try', *gebieden* 'order', *gelasten* 'order', *plegen* 'be used to', *dienen* 'have to', *vermogen* 'be able to', *vertellen* 'tell' (when used as *te*-selecting verb; it is not archaic when it selects

of verbs that can take a *met*-complement as alternative to a *te*-complement. This is illustrated for *beginnen* 'begin' and *ophouden* 'stop' in (67) and (68) respectively.

(67) a. Ik **begin *te*** werken.
I begin to work
b. Ik **begin *met*** werken.
I begin with work
'I begin working.'

(68) a. Ik **hou op *te*** werken.
I hold up to work
b. Ik **hou op *met*** werken.
I hold up with work
'I stop working.'

Table 5.1 indicates for each *te*-selecting verb whether it can appear in the following three constructions: (i) Verb Raising [VR], (ii) the Third Construction [3C], and (iii) Extraposition [EP]. All three types were already illustrated for *proberen* 'try' in (53), repeated here in (69) for convenience.

(69) a. ... dat ik ***hem*** heb ***proberen*** (* ***om***) te bellen. VR
... that I him have tried.IPP for to call
b. ... dat ik ***hem*** heb ***geprobeerd*** (* ***om***) te bellen. 3C
... that I him have tried.PTCP for to call
c. ... dat ik heb ***geprobeerd*** (***om***) ***hem*** te bellen. EP
... that I have tried.PTCP for him to call
'... that I have tried to call him.'

In the VR construction (69-a), the embedded object appears to the left of all clause final verbs (i.e. verb clustering has taken place), and the *te*-selecting verb appears in IPP form rather than as a past participle. In the Third Construction (3C) (69-b), the embedded object also appears to the left of all clause final verbs. However, the *te*-selecting verb appears in past participle form. In the EP construction (69-c), the embedded object appears after the *te*-selecting verb, which appears in past participle form.

Table 5.1 also indicates whether, when embedded under a perfective auxiliary, the *te*-selecting verb itself obligatory occurs in IPP form [yes], can occur either in IPP form or as a past participle [opt.], or can only occur as a past participle [no]. For four verbs, namely *blijken* 'turn out', *schijnen* 'seem', *lijken* 'seem', and *hebben* 'have to', the IPP property is indicated with 'NA', because these verbs cannot be embedded under a perfective auxiliary. Third, for each verb the table also indicates whether the presence of *te* is optional throughout the entire language area [opt.], only in a specific region [%opt.], obligatorily present [yes], or optional, but with a difference in meaning between the bare

a *dat*-complement), *verzuimen* 'fail'.

or *te*-infinitive [opt.].[9] Fourth, the table lists whether the verb can take an alternative *om te*-complement, a *met*-complement, and/or a *dat*-complement [yes], or not [no]. There are a number of verbs that can only take an alternative *om te*-complement when there is an expletive object *er* or *het* present in the clause. This is illustrated for *wagen* 'dare' and *ophouden* 'stop' in (70) and (71) respectively.

(70) a. ... dat ik **waag *te*** werken.
... that I dare to work

b. ... dat ik *(***het***) **waag *om*** te werken.
... that I it dare for to work
'... that I dare to work.'

(71) a. ... dat ik **ophou *te*** werken.
... that I stop to work

b. ... dat ik *(***er***) mee **ophou *om*** te werken.
... that I EXPL with stop for to work
'... that I stop working.'

Lastly, the table provides information about whether the complement can independently refer to the future (irrealis) or past [yes], or not [no], or whether this depends on which of the two distinct meanings of the verb is being used [opt.]. All judgments in the table are based on configurations in which these verbs are embedded under a perfective auxiliary.[10,11]

As can be seen from the top of Table 5.1, the only verbs that show optionality regarding *te*, the first eight verbs in the table, are the verbs I have listed in (55) as verbs that select a small complement. The first five verbs, *hoeven* 'need', *lopen* 'walk', *zitten* 'sit', *staan* 'stand' and *liggen* 'lie', have exactly the same profile. That is, they are the only verbs (i) for which the *te* is optional throughout the language area, (ii) which cannot take an alternative *om te*-, *met*- or *dat*-complement, and (iii) whose complement cannot independently refer to the future or past. In addition, they can only occur in the VR construction, and obligatorily occur in IPP form. The fact that these five verbs cannot take an alternative complement and that their complement can never refer to the future or past, I take to be strong indications that they select a small complement. In the next subsection, I show that the optionality of *te* is

[9] One verb has a different label in this category, namely *weten*$_1$ 'know where'. This verb always takes a *te*-infinitive in the Netherlands, whereas it takes a bare infinitive in Flanders, meaning that there is no optionality of *te* in either region, see Den Dikken and Zwart (1996). This is indicated with the label [yes$_{NL}$/no$_{VL}$].

[10] Unless the verb cannot be embedded under a perfective auxiliary, as is the case for *blijken*, *schijnen*, *lijken*, and *hebben*, see above.

[11] This comment is especially relevant when it comes to the optionality of *te* for the first eight verbs in the table. *Te* is optional (either throughout the entire language area or in a specific region) when the *te*-selecting verb is itself embedded under another verb. However, when the *te*-selecting verb is the finite verb of the clause (and thus not embedded under another verb), *te* becomes obligatory in all cases. I will come back to this finite/non-finite split in chapter 6.

Verb	VR	3C	EP	IPP	TE	OM	MET	DAT	irrealis	past
Hoeven 'need'	yes	no	no	yes	opt.	no	no	no	no	no
Lopen 'walk'	yes	no	no	yes	opt.	no	no	no	no	no
Zitten 'sit'	yes	no	no	yes	opt.	no	no	no	no	no
Staan 'stand'	yes	no	no	yes	opt.	no	no	no	no	no
Liggen 'lie'	yes	no	no	yes	opt.	no	no	no	no	no
Durven 'dare'	yes	yes	yes	opt.	opt.	no	no	no	no	no
Beginnen 'begin'	yes	yes	yes	opt.	yes (%opt)	no	yes	no	no	no
Proberen 'try'	yes	yes	yes	opt.	yes (%opt)	yes	no	no	opt.	no
Blijken 'turn out'	yes	no	no	NA	yes	no	no	yes	yes	yes
Lijken 'seem'	yes	no	no	NA	yes	no	no	yes	yes	yes
Schijnen 'seem'	yes	no	no	NA	yes	no	no	yes	yes	yes
Hebben 'have/have to'	yes	no	no	NA	opt.	no	no	no	no	no
Komen 'come'	yes	no	no	opt.	opt.	opt.	no	no	opt.	no
Weten$_1$ 'know where'	yes	no	no	yes	yes$_{NL}$/no$_{VL}$	no	no	no	no	no
Weten$_2$ 'manage'	yes	no	no	yes	yes	no	no	no	no	no
Zien 'see/try'	yes	no	no	yes	opt.	no	no	no	opt.	no
Horen 'hear/ought to'	yes	no	no	yes	opt.	no	no	no	opt.	no
Voelen 'feel/think'	yes	no	no	yes	opt.	no	no	no	opt.	no
Denken 'think'	yes	yes	yes	yes	yes	no	no	yes	yes	no
Helpen 'help'	yes	yes	yes	opt.	opt.	yes	yes	no	opt.	no
Leren 'teach'	yes	yes	yes	opt.	opt.	yes	no	yes	opt.	no
Menen 'think'	yes	yes	yes	no	yes	no	no	yes	yes	yes
Trachten 'try'	yes	yes	yes	opt.	yes	yes	no	no	yes	no
Wagen 'dare'	yes	yes	yes	no	yes	yes(*het*)	no	no	no	no
Weigeren 'refuse'	yes	yes	yes	opt.	yes	yes	no	no	yes	no
Wensen 'wish'	yes	yes	yes	no	yes	yes	no	yes	yes	yes
Aanraden 'recommend'	no	yes	yes	no	yes	yes	no	no	yes	no
Adviseren 'advise'	no	yes	yes	no	yes	yes	no	no	yes	no
Begeren 'desire'	no	yes	yes	no	yes	yes	no	no	yes	no
Beloven 'promise'	no	yes	yes	no	yes	yes	no	yes	yes	no
Beogen 'intend'	no	yes	yes	no	yes	yes	no	no	yes	no
Besluiten 'decide'	no	yes	yes	no	yes	yes	no	yes	yes	no
Bevelen 'order'	no	yes	yes	no	yes	yes	no	no	yes	no
Beweren 'claim'	no	yes	yes	no	yes	no	no	yes	yes	yes
Denken 'plan'	no	yes	yes	no	yes	no	no	yes	yes	no
Dreigen 'threaten'	no	yes	yes	no	yes	yes	no	no	yes	no
Dwingen 'force'	no	yes	yes	no	yes	yes	no	no	yes	no
Eisen 'demand'	no	yes	yes	no	yes	yes	no	yes	yes	no
Geloven 'believe'	no	yes	yes	no	yes	yes	no	yes	yes	yes
Hopen 'hope'	no	yes	yes	no	yes	yes	no	yes	yes	yes
Opdragen 'order'	no	yes	yes	no	yes	yes	no	no	yes	no
Verbieden 'forbid'	no	yes	yes	no	yes	yes	no	no	yes	no
Vergeten 'forget'	no	yes	yes	no	yes	yes	no	yes	yes	yes
Verklaren 'declare'	no	yes	yes	no	yes	no	no	yes	yes	yes
Verleren 'unlearn'	no	yes	yes	no	yes	yes(*het*)	no	no	no	no
Vermijden 'avoid'	no	yes	yes	no	yes	yes(*het*)	no	no	no	no
Verplichten 'oblige'	no	yes	yes	no	yes	yes	no	no	yes	no
Verwachten 'expect'	no	yes	yes	no	yes	yes	no	yes	yes	yes
Verzoeken 'request'	no	yes	yes	no	yes	yes	no	no	yes	no
Voorstellen 'propose'	no	yes	yes	no	yes	yes	no	yes	yes	no
Vragen 'ask'	no	yes	yes	no	yes	yes	no	no	yes	no
Vrezen 'fear'	no	yes	yes	no	yes	yes	no	yes	yes	yes
Zeggen 'say'	no	yes	yes	no	yes	no	no	yes	yes	yes
Aansporen 'urge'	no	no	yes	no	yes	yes	no	no	yes	no
Begrijpen 'understand'	no	no	yes	no	yes	no	no	yes	yes	yes
(Be)merken 'notice'	no	no	yes	no	yes	no	no	yes	yes	yes
Beseffen 'realise'	no	no	yes	no	yes	no	no	yes	yes	yes
Betreuren 'regret'	no	no	yes	no	yes	yes(*het*)	no	yes	yes	yes
Haten 'hate'	no	no	yes	no	yes	yes(*het*)	no	yes	yes	yes
Zich herinneren 'remember'	no	no	yes	no	yes	yes(*het*)	no	yes	yes	yes
Inzien 'realise'	no	no	yes	no	yes	no	no	yes	yes	yes
Lukken 'succeed'	no	no	yes	no	yes	yes(*het*)	no	no	no	no
Ontdekken 'discover'	no	no	yes	no	yes	no	no	yes	yes	yes
Opgeven 'give up'	no	no	yes	no	yes	yes(*het*)	no	no	no	no
Ophouden 'stop'	no	no	yes	no	yes	yes(*er*)	yes	no	no	no
Zich realiseren 'realise'	no	no	yes	no	yes	no	no	yes	yes	yes
Toestaan 'permit'	no	no	yes	no	yes	yes(*het*)	no	yes	yes	no
Uitnodigen 'invite'	no	no	yes	no	yes	yes	no	no	yes	no
Verafschuwen 'abort'	no	no	yes	no	yes	yes(*het*)	no	yes	yes	yes
Veronderstellen 'suppose'	no	no	yes	no	yes	no	no	yes	yes	yes
Verwijten 'reproach'	no	no	yes	no	yes	no	no	yes	no	yes

Table 5.1: Properties of *te*-selecting verbs

the result of the tight structural relationship between these *te*-selecting verbs and their complement.

Let us now look at the next three verbs in the table, *durven* 'dare', *beginnen* 'begin', and *proberen* 'try'. First, we can see that *durven* is the same as the five highest verbs in that it cannot select any alternative complement, and in that its complement cannot refer to the future/past. Furthermore, *te* is optional. The only difference between *durven* and the five highest verbs, is that *durven* can occur in all three constructions: Verb Raising, Third Construction, and Extraposition (72). The others can only occur in the Verb Raising construction.

(72) a. ...dat ik ***het boek*** niet heb ***durven*** **(*te*)** lezen. VR
...that I the book not have dare.IPP to read
b. ...dat ik ***het boek*** niet heb ***gedurfd*** ***(*te*)** lezen. 3C
...that I the book not have dare.PTCP to read
c. ...dat ik niet heb ***gedurfd*** (***om***) ***het boek*** ***(*te*)** lezen. EP
...that I not have dare.PTCP for the book to read
'...that I haven't dared to read the book.'

It is important to note here that *te* is only optional with *durven* in the Verb Raising construction (1-a), and remains obligatory in the other two constructions (1-b)–(1-c). This also holds for *te* selected by the next two verbs in the table, namely *beginnen* 'begin' and *proberen* 'try'. These two verbs allow optional *te* in the VR construction in most regions of Flanders, but *te* remains obligatory in the 3C and EP constructions (73)–(74).

(73) a. ...%dat ik ***het boek*** niet ben ***beginnen*** **(*te*)** lezen.
...that I the book not am begin.IPP to read
b. ...dat ik ***het boek*** niet ben ***begonnen*** ***(*te*)** lezen.
...that I the book not am begin.PTCP to read
c. ...dat ik niet ben ***begonnen*** ***het boek*** ***(*te*)** lezen.
...that I not have begin.PTCP the book to read
'...that I haven't begun to read the book.'

(74) a. ...%dat ik ***het boek*** niet heb ***proberen*** **(*te*)** lezen.
...that I the book not have try.IPP to read
b. ...dat ik ***het boek*** niet heb ***geprobeerd*** ***(*te*)** lezen.
...that I the book not have try.PTCP to read
c. ...dat ik niet heb ***geprobeerd*** (***om***) ***het boek*** ***(*te*)** lezen.
...that I not have try.PTCP for the book to read
'...that I haven't tried to read the book.'

Summing up, the optionality of *te* is uniformly attested with verbs that only occur in the VR construction, and never take an alternative type of complement (an *om te*-, *met*- or *dat*-complement). Verbs that behave like that are *hoeven*, *lopen*, *zitten*, *staan* and *liggen*. They select a very small complement, which cannot independently refer to the future or past. If a verb only partly allows optionality of *te* in its complement, it does so only in the VR construction, and not in the 3C or EP constructions. Verbs that behave like that are *durven*, *beginnen* and *proberen*. The contrast between the optionality of *te* in very small complements and its obligatory nature in larger complements, seems to suggest that we are dealing with two distinct types of *te* in Dutch. In other words, *te* seems to be a multifunctional morpheme, with its different uses corresponding to smaller and larger chunks of syntactic structure. In the next subsection, I discuss this idea in more detail.

5.3.3 *Te* as a multifunctional morpheme

In this subsection I argue that *te* is a multifunctional morpheme, whose different uses correspond to different syntactic sizes. I propose that we have to distinguish between the following three *te*'s:

1. *te* as a preposition
2. *te* spelling out a functional head in a non-finite clause, henceforth 'head-*te*'
3. *te* spelling out an uninterpretable but valued [T]-feature on v, henceforth 'feature-*te*'

An example of each type given in (75)–(77) respectively.

(75) Ik ben geboren **te** 's-Hertogenbosch.
I am born in 's-Hertogenbosch
'I was born in 's-Hertogenbosch.'

(76) a. Ik zeg naar huis **te** zijn gegaan.
I say to home to be go
'I say I went home.'
b. Ik besluit naar huis **te** gaan.
I decide to home to go
'I decide to go home.'

(77) Ik heb niet hoeven **te** werken.
I have not need to work
'I didn't need to work.'

Prepositional *te* is embedded in a prepositional structure. As this chapter is not concerned with the prepositional use of *te*, however, I leave the exact nature of this use out of the discussion, and focus on the syntactic structure of head-*te* and feature-*te*. I propose the following. Head-*te* spells out a functional head in the syntactic superstructure of the verb, whereas feature-*te* is the spell out of a uninterpretable but valued T-feature on a little v-head.[12]

[12]Note that for German *zu* both types of analyses exist; some authors take *zu* to be an independent syntactic head (Sternefeld 1990, Stechow 1990, Hinterhölzl 2009), whereas others

Before presenting the syntactic structures of these two different *te*'s, I need to specify my take on Agree. I adopt a *Reverse Agree* approach to Agree, following Bjorkman (2011), Wurmbrand (2012b,a) et seq, and Zeijlstra (2012).[13] Wurmbrand (2012a)'s definition of Reverse Agree is given in (78).

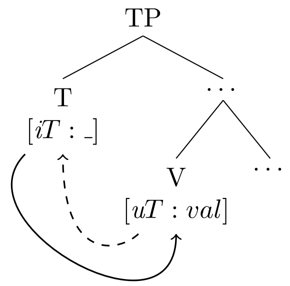
(78) *Reverse Agree*
A feature F:_ on α is valued by a feature F:val on β, iff
i) β asymmetrically c-commands α AND
ii) There is no γ, γ distinct from β, with a valued interpretable feature F such that γ commands α and is c-commanded by β.
(Wurmbrand 2012a:135)

I furthermore adopt Wurmbrand (2012a)'s approach to the distribution of verbal features. She assumes the four-way split among interpretable, uninterpretable, valued and unvalued features that was proposed by Pesetsky and Torrego (2007) and Bošković (2009). According to these authors, interpretability does not correlate with valuation, which means that all four combinations of (un)interpretability and (un)valuation are possible: *i*F:val, *i*F:_, *u*F:val and *u*F:_. In Reverse Agree, valuation is downward rather than upward, i.e. the valued feature c-commands (and gives its value to) the unvalued feature. In Standard Agree (Chomsky 2000, 2001) the verb enters the syntactic derivation with an uninterpretable but valued T(ense)-feature. T starts out with an unvalued interpretable T-feature. This feature probes down, and Agrees with the valued uninterpretable T-feature on V. As a result, it gets valued. This illustrated in (79). The black arrow indicates the direction of probing and Agree, the dotted arrow the direction of valuation. In the Reverse Agree approach, however, the direction of Agree and valuation are the exact opposite. The verb enters the derivation with an unvalued, uninterpretable T-feature. This feature probes up and Agrees with the valued [*i*T]-feature on T. As a result, the unvalued [*u*T]-feature on V gets valued. This is illustrated in (80), again with the black arrow indicating the direction of probing and Agree, and with the dotted arrow indicating the direction of valuation.

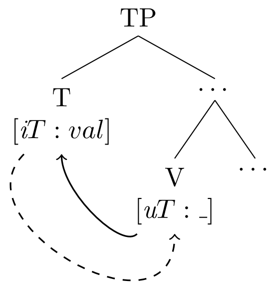
analyse it as a feature on the non-finite complement (Bader 1995, Meurers 2000, Vogel 2009). See Salzmann (2016:420-423) for more discussion. However, as far as I am aware, no one has thus far proposed that both analyses are on the right track, albeit with respect to different uses of *zu*, which is what I am proposing here for its Dutch cognate.

[13] I am not committed to Reverse Agree being the only way to Agree. However, I think verbal feature valuation (in Germanic) is most likely the result of Reverse Agree. See Bjorkman and Zeijlstra (2019, 2014), Carstens (2015), Preminger (2015), Smith (2015) for different views and detailed discussion of the possibility of having both Standard and Reverse Agree or only one of the two as part of the language faculty.

(79) **Standard Agree**



(80) **Reverse Agree**

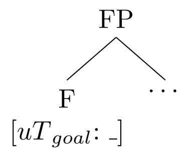


With the four-way split of (un)interpretability and (un)valuation, the question arises how we know whether the interpretable or the uninterpretable feature is the valued one and which is the unvalued one. Pesetsky and Torrego (2007:227) suggest that the T-feature on T is interpretable but unvalued, since according to them T learns its value in finite clauses from the finite verb. However, Wurmbrand (2012a) rightly notes that it can also be the other way around: the verb learns its tense value from T. I follow Wurmbrand (2012a) in the feature specifications she proposes for verbal heads (in Germanic). These assumptions are the following. First, all functional clausal heads (i.e. T, Aux, Mod, Asp et cetera) have a (typically) valued [*i*T]-feature, with its value corresponding to the semantic value of the head (e.g. past, modal, perfect et cetera). Second, all verbal heads (i.e. V, but also Aux, Mod and Asp) have a [*u*T]-feature, which is (typically) unvalued.[14] Third, at PF the *u*T-feature on V is what is morphologically realised (see also Stechow (2003) et seq. and Grønn and Stechow (2011)). So, for example, when in English a modal values the [*u*T:_] on a verb, this verb is realised as an infinitive, whereas when this feature is valued by a perfective auxiliary or a passive auxiliary, the verb is realised as a participle. Given that the V head is the only head in the functional sequence that is verbal but not functional, it is the only head that only has an unvalued *u*T-feature. All other clausal functional heads that are lexicalised by verbs, e.g. Aux, Mod, Asp et cetera, come with both a valued *i*T-feature (valued for their semantic value as perfect, modal, aspectual/progressive et cetera) and an unvalued *u*T-feature. I also adopt these assumptions, but diverge from Wurmbrand's approach on one point: I do not assume that the unvalued *u*T is on V, but on *v*. I make this assumption beacuse I take a lexical verb to consist (at least) of a root which is verbalised by a little *v* head, and thus do not assume there is such a thing as a V position in which the lexical verb is Merged. In my adapted approach,

[14] Wurmbrand (2012a) adds "typically" to these definitions because she wants to leave room for a certain amount of cross-linguistic variation with respect to this. The interested reader is referred to her paper for illustration and argumentation.

then, it is the value that uT on v receives that is morphologically realised on the verb after spell out.
With these ingredients of the analysis in place, let us now return to my proposal for two different types of *te* in the verbal domain: a clause marker and a verbal marker. head-*te* enters the derivation as a functional head, and comes with a T_{goal}-feature, which is unvalued.[15] The syntactic structure of head-*te* is given in (81).

(81) **head-*te***



The reason I assume the existence of a T_{goal}-feature on the functional head F which can be spelled out as *te* is based on the diachronic development of *te*. As already mentioned above, IJbema (2001) shows that *te* started out as a preposition. More specifically, it started out as a preposition with a spatial directional meaning, which indicates the goal that is intended to be reached. For example, *te 's-Hertogenbosch* in (82) would mean 'to(wards) 's-Hertogenbosch', with 's-Hertogenbosch being the goal of a journey, whereas nowadays it only has a spatial locative meaning, 'in 's-Hertogenbosch'.[16]

(82) Ik ga te 's-Hertogenbosch.
I go to 's-Hertogenbosch
'I'm going to 's-Hertogenbosch.'

The goal-oriented meaning of *te* was later extended to its use in the verbal domain. That is, when *te* would mark a verb, this would indicate a *temporal directive*, with the embedded verb as the goal of the matrix verb. For example, in (83), the goal of the matrix verb is that the embedded verb is executed, namely for Anna to leave.

(83) Anna besluit [te vertrekken].
Anna decides to leave
'Anna decides to leave.'

I implement this temporal directive/goal meaning by means of a T_{goal}-feature on the F-head that *te* is spelling out. This feature is furthermore uninterpretable: from the semantics of *te* itself we cannot deduce any temporal information. Thus, the functional head F of the clausal marker *te* enters the

[15] Note that it is very well possible that *te* has other features except for this T_{goal}-feature, but I refrain here from determining the exacT-feature bundle the head-*te* comes equipped with, as the T_{goal}-feature is the only one that is relevant for the purposes of this chapter.
[16] The spatial directive meaning of *te* is retained in certain fixed expressions in Dutch, like *te water gaan* 'to go into the water'.

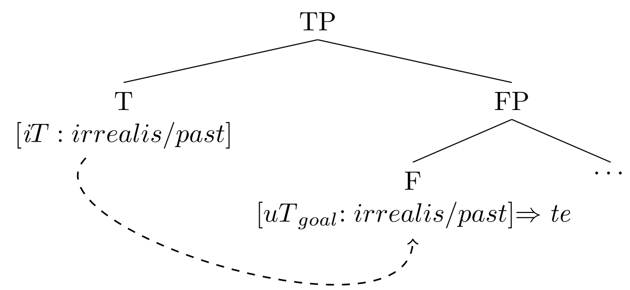
derivation with an unvalued T_{goal}-feature. Regarding its valuation, I propose the following. For the functional head to be spelled out as *te*, the *u*T-feature on F should be valued as either *irrealis* or *past*. I base this assumption on the fact that the *te*-complements selected by the verbs in (56) can independently refer to the future (irrealis) or to the past. An example of an irrealis *te*-complement is given in (84) and one of a past *te*-complement in (85).

(84) . . . dat Eva zegt de hele dag [$_{irrealis}$ te willen werken].
. . . that Eva says the entire day to want work
'. . . that Eva says she wants to work the entire day.'

(85) . . . dat Eva zegt de hele dag [$_{past}$ te hebben gewerkt].
. . . that Eva says the entire day to have worked
'. . . that Eva says to have worked the entire day.'

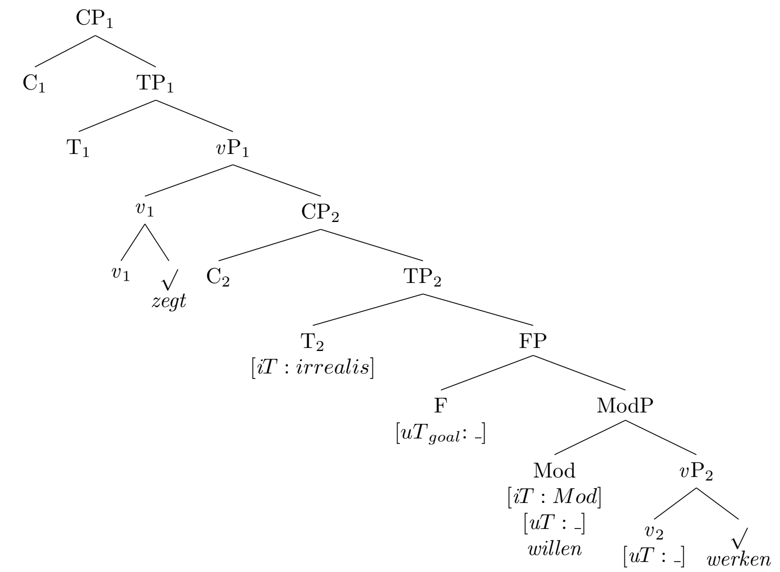
In other words, a verb that selects a large *te*-complement selects a TP complement in which T either has an [*i*T:irrealis]-feature or an [*i*T:past]-feature. Given that the complement indicates a temporal goal that lies either in the future (irrealis), or that has been accomplished already (past), the structure also contains a functional projection FP, the head of which is endowed with an unvalued uT_{goal}-feature. This feature gets valued by the interpretable T-feature on T. When it is valued as either irrealis or past, F gets spelled out as *te*, as illustrated in (86). I label this use of *te* 'head-*te*', as it marks the embedded clause as a (not yet/already realised) goal of the matrix verb.

(86) **Spell out of clausal marker *te***

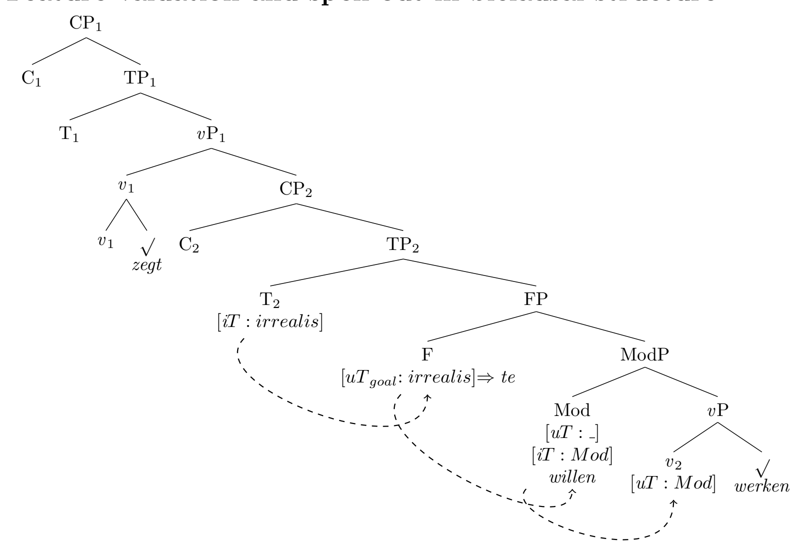


Structures in which the head-*te* surfaces are thus always biclausal. The verbs of the type in (56) select a complement that is at least as large as a TP. Head-*te* marks the embedded clause as a goal of the higher matrix verb. A simplified syntactic structure of (84) is given as an illustration in (87), with the feature valuation in that structure in (88).

(87) **Biclausal structure of *zeggen* and its *te*-complement**



(88) **Feature valuation and spell-out in biclausal structure**



The unvalued uT-feature on v is valued by the [iT:Mod]-feature on Mod, which results in the lexical verb *werken* 'work' being spelled out as a bare infinitive. The unvalued $u\mathrm{T}_{goal}$-feature on F is valued by the [iT:irrealis]-feature on T, which results in the spell out of *te*. The resulting embedded clause is thus: *te willen*.INF *werken*.INF. Note that, as also specified above, the Mod head comes with both an unvalued *u*T-feature and an [*i*T:Mod]-feature. The latter values the unvalued feature on *v*. Since I adopt a Reverse Agree approach, I thus assume that the *u*T-feature on Mod is valued by a higher head with a valued *i*T-feature. The next highest head with such a feature in (87) is T: it bears an [*i*T:irrealis]-feature. the *u*T-feature on Mod will thus be valued as irrealis. Anticipating the discussion of feature-*te* below, I will argue that this *te* is the spell out of a *u*T-feature on *v* if this feature is valued for *irrealis* or *past*. This begs the question of whether this [*u*T:irrealis]-feature on Mod results in Mod being spelled out as a *te*-infinitive. The answer is negative. The reason is that *te*, when it does not have its own functional projection in the syntactic structure, can only spell out this [*u*T:irrealis]-feature when it is specified on *v*. In other words, *v* is a position on which verbal morphology can be spelled out. Given that functional verbal heads, like Aux and Mod, lack such a *v* head, they are expected never to be able to bear verbal morphology when they are part of the functional sequence of a lexical verb. This is indeed the case: they can never bear perfective morphology (89) (i.e. the IPP-effect, see section 2.3.2), nor can they ever bear a verbal prefix like *be-*, *ver-*, *ont-* when they are in the functional sequence of a lexical verb (90). Both types of verbal morphology are only possible on lexical verbs (91). As I take a lexical verb to be a root verbalised by a *v* head, a logical position for this type of verbal morphology is *v*.

(89) a. . . . dat ik het boek heb ***gekund**/ **kunnen** lezen.
. . . that I the book have can.PTCP/ can.IPP read
'. . . that I have been able to read the book.'

b. . . . dat ik het boek heb ***gemoeten**/ **moeten** lezen.
. . . that I the book have must.PTCP/ must.IPP read
'. . . that I had to read the book.'

(90) a. . . . dat ik het boek heb ***be-kund**/ ***ver-kund**/ ***ont-kund**/
. . . that I the book have BE-can/ VER-can/ ONT-can
kunnen lezen.
can.IPP read

b. . . . dat ik het boek heb ***be-moeten**/ ***ver-moeten**/
. . . that I the book have BE-must/ VER-must/
***ont-moeten**/ **kunnen** lezen.
ONT-must can.IPP read

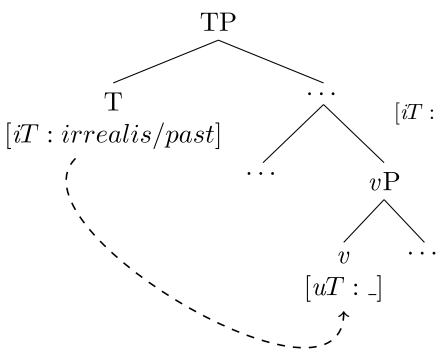
(91) a. . . . dat ik heb **ver-leerd**/ **ont-leerd** viool te spelen.
. . . that I have VER-learn/ ONT-learn violin to play
'. . . that I cannot play the violin any more.'

b. Zij is erg **beleerd**.
Zij is very BE-learned
'She is very knowledgeable.'

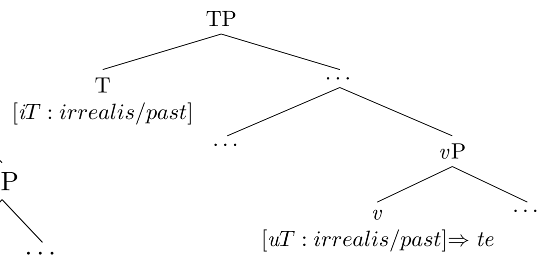
Returning to our structure in (88), the [uT:irrealis]-feature on Mod will not result in the spell out of *te* on the modal, as there is no v head for *te* to be realised on. The modal will thus receive a default spell out, which I take to be the bare infinitive. Summing up, based on my assumptions of Reverse Agree and the featural make-up of the verbal heads in a clause, the feature valuation in the structure in (88) will result in the sentence: *…zegt te willen werken* 'says to want.INF work.INF'.

Let us now move on to the feature-*te*. I want to propose that the feature-*te* is not an independent functional head in narrow syntax, but rather the spell out an uninterpretable T-feature on v, when this feature is valued for either *irrealis* or *past*. Feature-*te* is what we see in structures involving one of the verbs in (55), i.e. verbs that select a small complement. As we have seen above, the complement of verbs like *hoeven* and *zitten* cannot independently refer to the future or past, meaning that these verbs share the same clause with the complement verb, which is the lexical verb of the clause. I present the exact structures of these monoclausal configurations below, but first I abstractly show the Agree relation takes place which leads to the uT-feature on little v being spelled out as the feature-*te*. The Agree-relation is illustrated in (92) and resulting valuation and spell-out in (93).

(92) **Agree prior to spell out**

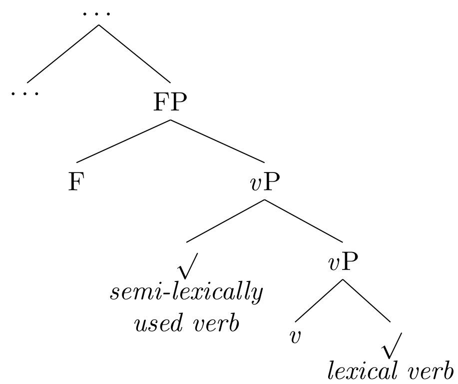


(93) **Feature-*te* spell out**

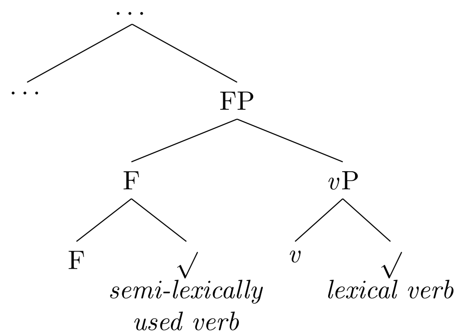


In chapter 2, I have argued that we need to distinguish between two stages of semi-lexical restructuring. The abstract structures are repeated here in (94) and (95).

(94) **Semi-lexical restructuring stage I**

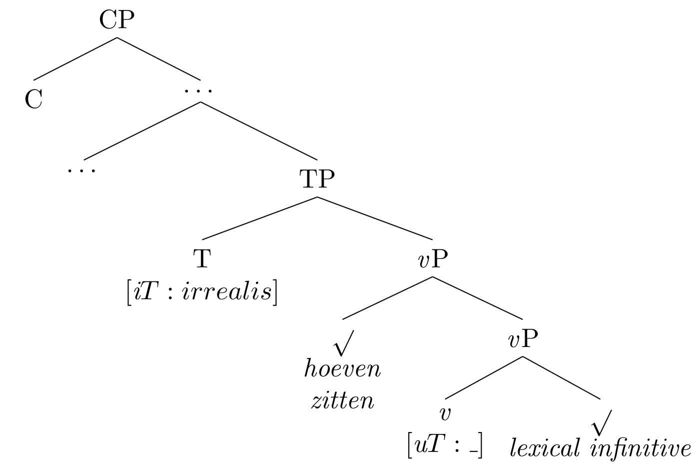


(95) **Semi-lexical restructuring stage II**



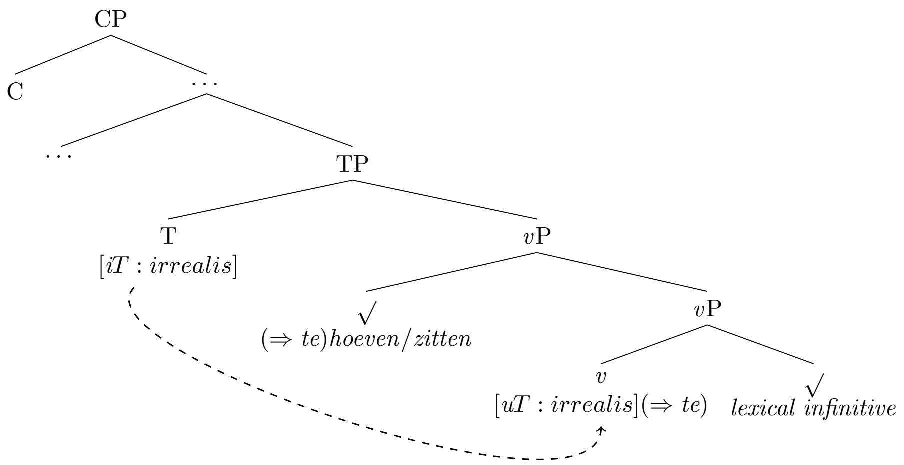
In the first stage of semi-lexical restructuring (59), the root of the semi-lexically used verb is merged above the lexical root, and is categorised as verbal material by the same v as the lexical root. In the second stage of semi-lexical restructuring (60), the semi-lexically used verb is merged with its functional head directly, in a separate workspace, and merged into the functional sequence of the lexical verb. Recall from section 5.2 that verbs *hoeven* and *zitten* were diagnosed as semi-lexically used verbs, in the case of *hoeven* even with an almost functional status. I therefore propose that *zitten* when used semi-lexically is always a case of semi-lexical restructuring of stage I, whereas *hoeven* is in the process of grammaticalisation from the first stage of semi-lexicality to the second stage. Returning now to the spell out of the feature-*te* in constructions in which *zitten* or *hoeven* selects a *te*-complement, I introduce the monoclausal structures of their configurations below. First, in (96) the structure prior to valuation and spell-out is given for the first stage of semi-lexicality, in which *zitten* uniformly falls, and *hoeven* partly.

(96) **Monoclausal structure of *hoeven/zitten* and its complement (semi-lexical stage I)**



The lexical verb is verbalised by a v head, which bears an unvalued uT-feature. This verbalising head projects a second time when the semi-lexically used root is Merged. Given that the semi-lexically used root is featureless, it is necessarily the verbalising head that projects, as the semi-lexically used root has no feature that can project. Given that the semi-lexically used and lexical verb constitute a monoclausal structure, there is only one TP in the structure. When v probes upward to look for an interpretable T-feature that can value it, it finds the valued feature on T. In case the iT-feature on T is valued for *irrealis* or *past*, and thus values the uT on v as either of those two values, at spell-out v is realised as *te*. However, v projects twice, once while Merging with the lexical root and once while Merging with the semi-lexically used root, As a result, a potential spell-out conflict arises. Given that the [uT:*irrealis/past*]-feature is technically present on v and all of its projections, it can be spelled out on the semi-lexically used root, on the lexical root, or both. As an illustration, the feature valuation and spell-out of this structure is given in (97). I propose that speakers can differ in which strategy they use to resolve this spell-out conflict. Some speakers will try to spell out the feature either on the highest projection, or on the lowest, or on both. For other speakers the spell-out conflict will lead to the feature just not being realized morphophonologically at all. In other words, we expect variation in the placement and presence of *te* in such a configuration (i.e. the different *te*-placement options and *te*-drop). In subsection 5.4.3, I work this out in more detail.

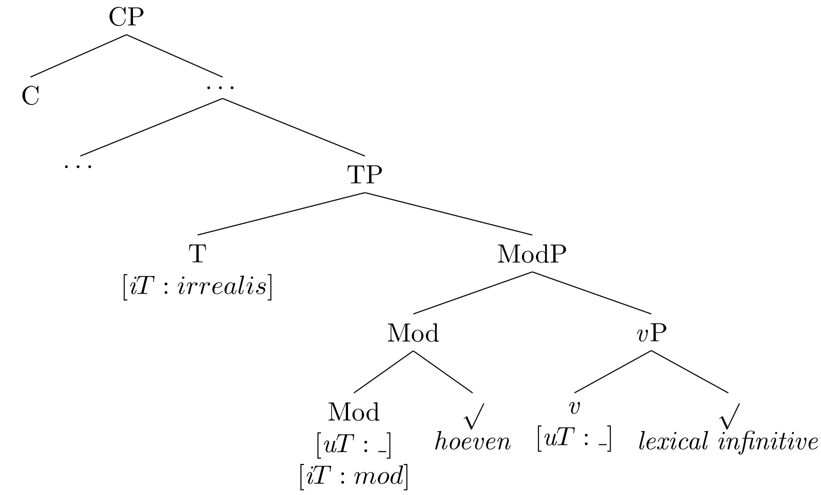
(97) **Feature valuation and spell-out in monoclausal structure (semi-lexical stage I)**



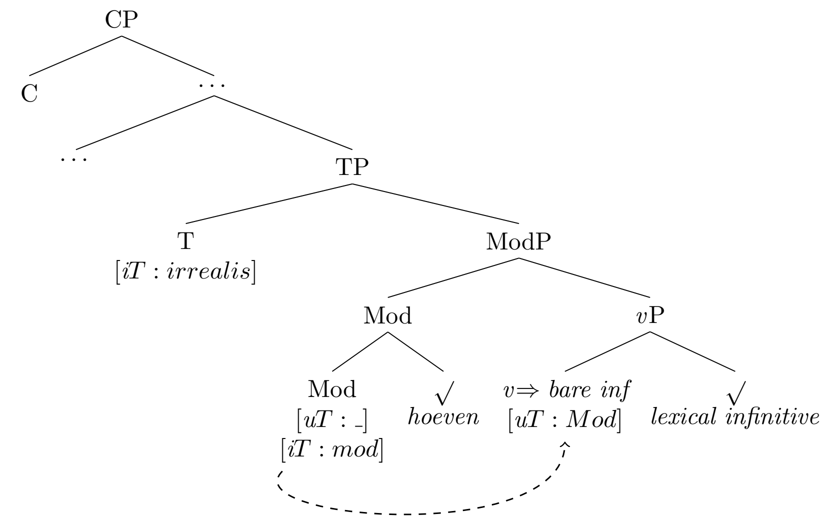
It is important to note that strictly speaking these semi-lexically used verbs no longer c-select a *te*-infinitive. Whereas the verbs that select a TP/CP-sized *te*-complement can really be seen as c-selecting for this type of complement (see e.g. the structure in (87)), verbs like *hoeven* and *zitten* are themselves (a low) part of the functional superstructure of a lexical verb. The *te* that surfaces on the lexical verb (or on the semi-lexically used verb in case the [*u*T:irrealis]-feature is spelled out on the highest *v* projection, i.e. in cases of *te*-raising) is not selected by *hoeven/zitten*, but is a 'side effect' of the tense of the clause being specified as irrealis or past.

Let us now look at the structure of stage II of semi-lexical restructuring. As argued above, for some speakers this is the underlying structure when *hoeven* combines with an embedded verb. As was illustrated in (60), in the second stage of semi-lexicality, the root of the semi-lexically used verb is Merged with a functional head, and the two are then Merged as a complex head in the functional structure of the lexical verb. In the case of *hoeven*, the functional head it combines with is Mod; modal *hoeven* expresses necessity (and more specifically the absence of necessity, given *hoeven*'s NPI nature). The structure of *hoeven* combined with a lexical verb in the second stage of semi-lexicality is given in (98). In this stage of semi-lexicality, *v* only verbalises the lexical verb. As was already discussed above, functional heads in the verbal structure come with both an interpretable T-feature valued for their semantic value, and an unvalued uninterpretable T-feature. This is also the case for the Mod head that Merges with the semi-lexically used root in (98). As was also already discussed above, however, whatever value this *u*T-feature on Mod receives, it will always be spelled out as an infinitive, because verbal morphology needs a v head to be realised. When the unvalued uT-feature on v probes up, the first interpretable T-feature it finds is the one on Mod. It will therefore be valued as Mod, and be spelled out as a bare infinitive. The feature valuation and spell-out is illustrated in two steps in (99) and (100).

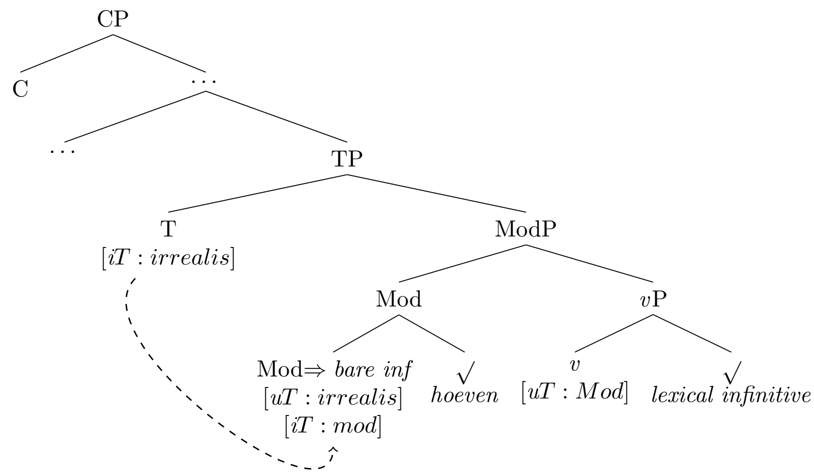
(98) **Monoclausal structure of *hoeven* and the lexical verb (semi-lexical stage II)**



(99) **Feature valuation and spell-out monoclausal structure step one (semi-lexical stage II)**



(100) **Feature valuation and spell-out monoclausal structure step two (semi-lexical stage II)**



We thus see that in this stage of semi-lexicality, the [*i*T:Mod]-feature on the Modal head which combines with the semi-lexically used root is the feature that values the uninterpretable T-feature on *v*, which means that the lexical verb will always be spelled out as a bare infinitive, and not as a *te*-infinitive. *Te* can also not be spelled out on the semi-lexically used verb itself, as there is no *v* projection that Merges with the semi-lexically used root. In this type of configuration, *te* is thus always absent (i.e. *te*-drop).

With all of my assumptions now illustrated and laid out, let us move on to the next subsection in which I show how my proposal deals with the four phenomena that we uncovered in the questionnaire data, namely *te*-raising, *te*-lowering, *te*-doubling and *te*-drop, as well as *te*-in-situ, for each type of verb cluster that was tested in the questionnaire study.

5.4 The analysis of the phenomena

5.4.1 *Te*-placement and -presence in cluster type Ia

We start with cluster type Ia, which is repeated again here in (101).

(101) Anne **zegt** op haar comfortabele stoel [*te* willen$_1$ blijven$_2$
Anne says on her comfortable chair to want.INF remain.INF
zitten$_3$].
sit.INF

'Anne says she wants to remain seated on her comfortable chair.'
CLUSTER TYPE IA

Recall from subsection 4.3.7 that in this cluster type: (i) *te*-in-situ is accepted by all speakers, (ii) *te*-raising is excluded because *te* is already situated on the highest verb in the cluster (i.e. V_1 *willen*), (iii) *te*-lowering is infrequent, and *te* mainly lowers to V_2 (>80% of the lowering cases) rather than V_3, and (iv) *te*-doubling is also infrequent, with the most frequent configuration being V1-*te*-V2-*te*-V3 (>55% of the doubling cases). Furthermore, *te*-drop is seen as noise in this cluster type per the discussion in the previous subsection.

Before I can show how my proposal accounts for these facts, I first need to establish the status (functional or semi-lexical) of V1 and V2 in the cluster. V3 is the lexical verb of the cluster, and thus unquestionably has a lexical status. V1 *willen* 'want' is a modal verb with a functional status: it always occurs in IPP form when embedded under a perfective auxiliary (102), it rejects extraposition of the embedded infinitive (144), and does not impose animacy restrictions on the subject (104).

(102) ...dat hij heeft **willen**/ ***gewild** gaan.
...that he has want.IPP/ want.PTCP go
'...that he wanted to go.'

(103) ...*dat hij heeft **willen**/ **gewild** [een boek kopen.]
...that he has want.IPP/ want.PTCP a book buy

(104) ...dat ***het*** maar niet heeft **willen** sneeuwen deze winter.
...that it PRT not has want.IPP snow this winter
'...that it didn't snow at all this winter.'

V2 *blijven* 'remain' is an aspectual verb with functional status, as it always occurs in IPP form when embedded under a perfective auxiliary (105), rejects extraposition (106), and does not impose animacy restrictions on the subject (107).

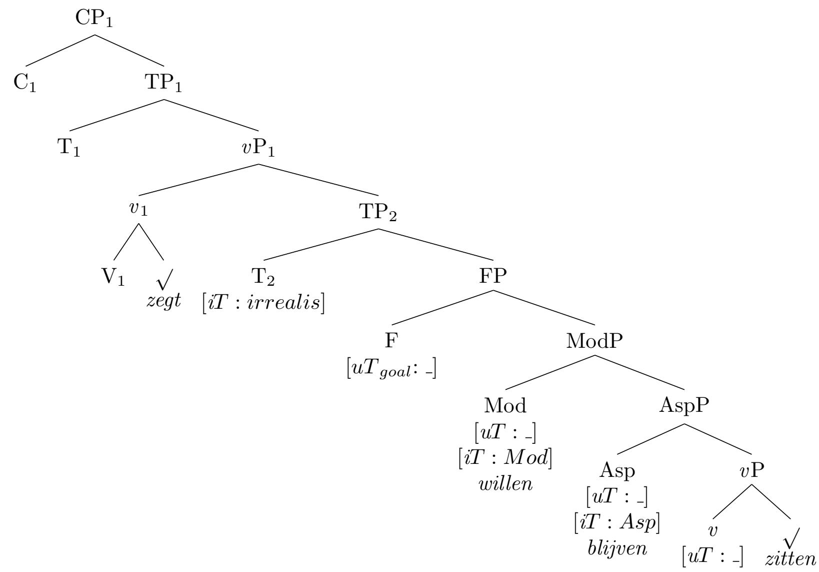
(105) ...dat hij is **blijven**/ ***gebleven** eten.
...that he is remain.IPP/ remain.PTCP eten
'...that he stayed over for dinner.'

(106) ...*dat hij is **blijven**/ **gebleven** [de oefening herhalen].
...that he is remain.IPP/ remain.PTCP the exercise repeat
'...that he kept on repeating the exercise.'

(107) ...dat ***het*** is **blijven** sneeuwen.
...that it is remain.IPP snow
'...that it continued to snow.'

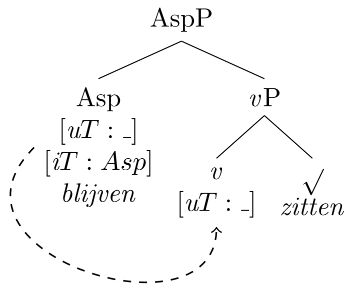
The verb that selects the *te*-infinitive is *zeggen* 'say'. This verb selects a complement that is at least as large as TP. In the test sentence in (101), the TP complement is an irrealis one: the verb cluster *willen blijven zitten* 'want to remain seated' has not been realised yet. This means that the iT-feature on the embedded TP is valued for *irrealis*. As the complement furthermore is the directive/goal, there is a FP below the embedded TP the head of which bears an unvalued $u\mathrm{T}_{goal}$-feature. Below this functional projection the three verbs of the verb cluster follow: V1 *willen* 'want' in Mod, V2 *blijven* 'remain' in Asp, and the lexical root, verbalised by a verbalising head, which bears an unvalued uT-feature. The structure of the sentence is given in (108). All irrelevant projections (as well as any arguments) are left out for ease of exposition.

(108) **Cluster type Ia: structure**



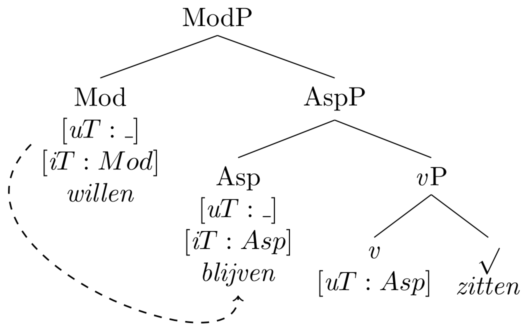
The valuation of the relevant T-features on the functional and verbal heads in the embedded complement clause are given step by step in (109)–(116).

(109) **Cluster type Ia: valuation step I**



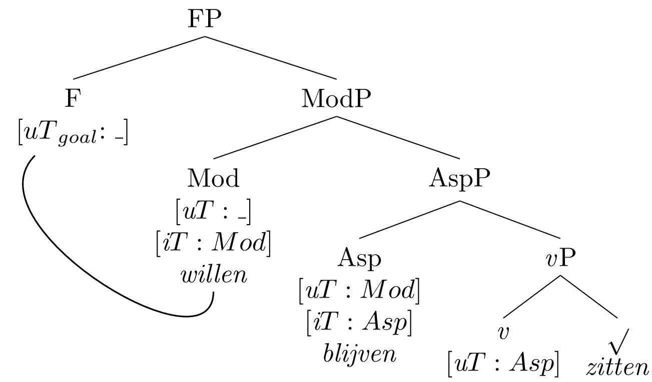
In (109), the uT-feature on v gets valued for *Asp* by [iT:Asp] on Asp.

(110) **Cluster type Ia: valuation step II**



In (110), the *u*T-feature on Asp gets valued for *Mod* by the [*i*T:Mod]-feature on Mod.

(111) **Cluster type Ia: valuation step III**

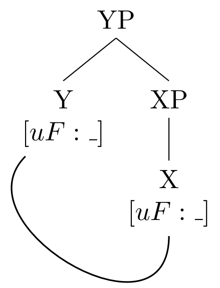


In (111), the unvalued *u*T-feature on Mod probes up, but only finds an unvalued $u\mathrm{T}_{goal}$-feature on F. As this feature cannot value the *u*T-feature on Mod, the two features Agree without valuation taking place. The idea that two unvalued features can establish an Agree relation has been proposed by Pesetsky and Torrego (2007). They assume a feature sharing version of Agree; their definition repeated here in (112).

(112) *Agree* (Feature sharing version)
(i) An unvalued feature F (a *probe*) on a head H at syntactic location α (F_α) scans its c-command domain for another instance of F (a *goal*) at location β (F_β) with which to agree.
(ii) Replace F_α with F_β, so that the same feature is present in both locations (Pesetsky and Torrego 2007:268)

They refer to the distincT-features that undergo Agree with each other as 'occurrences of F'. The two features that are linked after Agree – i.e. they form a feature pair, as they share the same feature – are called 'instances of F'. In other words, Agree turns two occurrences of F into two instances of F. Given their definition of Agree, it is possible that an unvalued occurrence of F Agrees with another unvalued occurrence of F, resulting in two instances of the same unvalued feature. When one of the instances of F in a later Agree operation becomes valued, the other instance of that F is valued as well. Since they use the four-way split created by the (un)interpretability and the (un)valued nature of features, they also assume that Agree between two uninterpretable features is possible, just as long as at some point one of the instances of this feature Agrees with an interpretable feature as well (Pesetsky and Torrego 2007:272).[17] I follow Pesetsky and Torrego (2007) in assuming that uninterpretable, unvalued features can enter an Agree relation with each other, and will then be linked. However, I do not take the nature of this link to be an actual feature *sharing* link, but rather a connection between the two occurrences of the features, which, when one of those features becomes valued, also results in the valuation of the other feature to which it is connected. An abstract illustration of the connection that is thus established is given in (113), and the parasitic valuation of the connected features in (114).

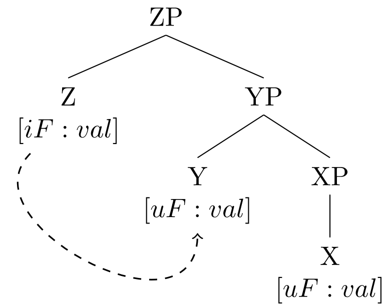
(113) **Agree establishes a connection between two unvalued F's**



In other words, when an unvalued feature F probes for a goal F, and the first goal F is an unvalued occurrence of this F, they will Agree and be connected (but not replaced). As soon as one of the connected features gets valued, the other feature receives the same valuation. One of the unvalued features can thus be seen as a sort of parasite on the other, waiting for valuation to take place.

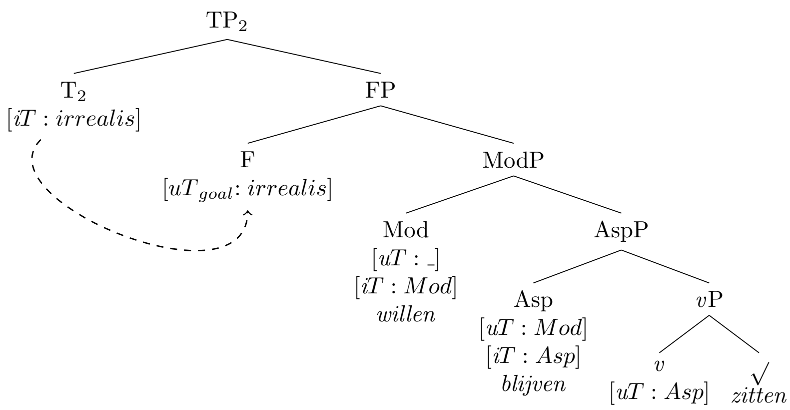
[17]See also Haegeman and Lohndal (2010) for the possibility of Agree taking place between two uninterpretable features.

(114) **Parasitic valuation of the connected F's**

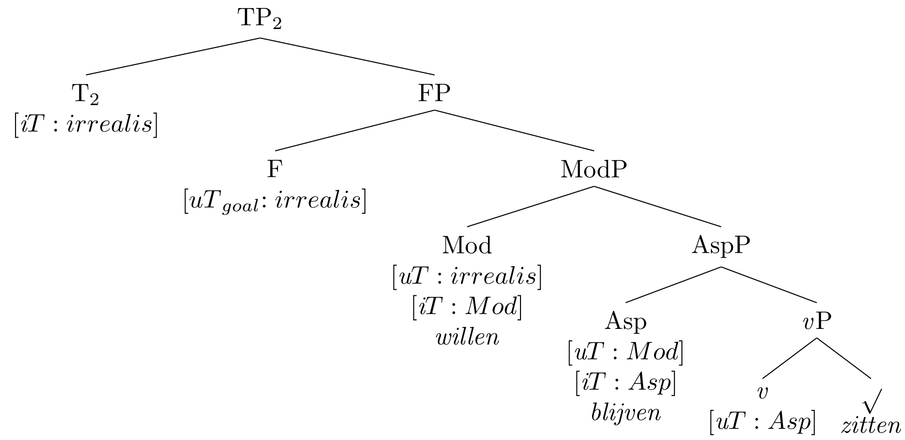


After this brief detour on valuation of two connected, unvalued features, let us return to the derivation and feature valuation in cluster type Ia. As just argued for, in (111), the unvalued uT-feature on Mod Agrees with the unvalued $u\mathrm{T}_{goal}$-feature on F, establishing a connection between the two features. After that, T is Merged, bearing a valued iT-feature. The unvalued $u\mathrm{T}_{goal}$-feature on F can now be valued by the feature on T (115), after which the connected uT-feature on Mod gets parasitically valued as well (116). The spell out of the cluster is given in (117).

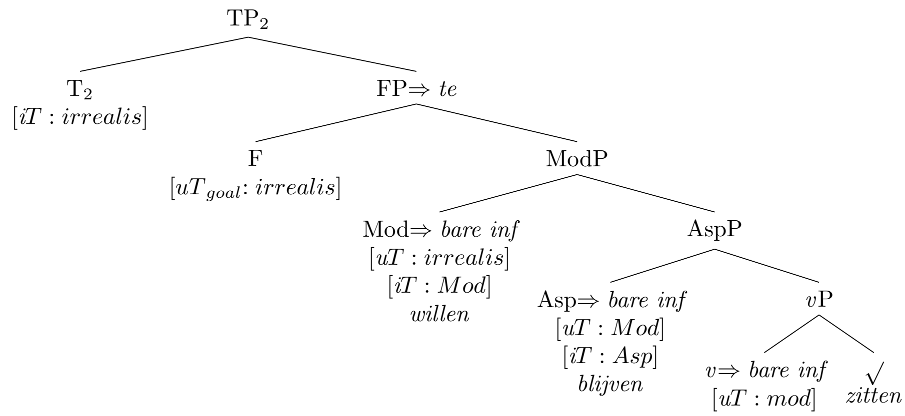
(115) **Cluster type Ia: valuation step IVa**



(116) **Cluster type Ia: valuation step IVb**



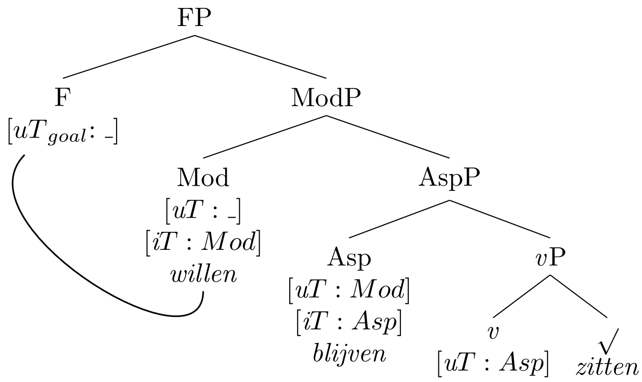
(117) **Cluster type Ia: spell out**



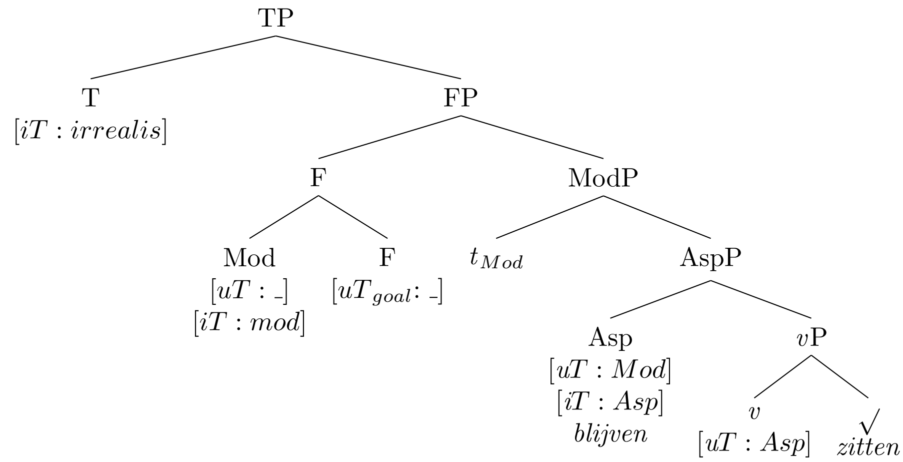
The [*u*T:irrealis]-feature on F is spelled out as *te*. Even though Mod bears the same feature, *te* cannot be spelled out on the Mod head, as it needs either a separate functional head (F) or a *v* head to be hosted on. Mod is thus spelled out as a bare infinitive. The [*u*T:Mod]-feature on Asp results in Asp being spelled out as a bare infinitive as well. The same holds for the [*u*T:Mod]-feature on *v*. Summing up, there is an F head which is spelled out as *te* at the beginning of the verb cluster, followed by three bare infinitives: *te-willen*$_1$-*blijven*$_2$-*zitten*$_3$. The standard way in which the derivation and feature evaluation of this cluster happens, thus results in the 'correct' placement of *te* in this cluster: on V1 *willen* 'want'. Recall that *te*-drop is virtually unattested in this cluster, and therefore taken to be noise. As there is no reason why Agree between the [*i*T:irrealis]-feature on T and the [*u*T$_{goal}$:_]-feature on F would fail, the nonexistence of *te*-drop in this cluster type is what we would expect based on the current analysis.

Let us now move on to *te*-lowering in this cluster. Recall that it is a rather infrequent phenomenon, and that *te* almost always lowers to V2, rather than V3. For the analysis of *te*-lowering, I propose the following. At the point in the derivation in which the uT on Mod probes up and establishes an Agree relation and thus a connection with the unvalued $u\mathrm{T}_{goal}$-feature on F (118), for some speakers this connection between the two features results in head movement of Mod to F (119). At the next step, F probes up and finds the valued iT-feature on T. The uninterpretable T_{goal}-feature gets valued for *irrealis* by the [iT:irrealis]-feature on T, and the same happens for the uT-feature on Mod, in a parasitic way (120). The spell-out of the cluster is given in (121). The Mod-F complex is spelled out as *willen-te*, followed by the other two verbs spelled out as bare infintives. The entire cluster is thus spelled out as: *willen-te-blijven-zitten*, i.e. *te*-lowering onto V2. Note that it is now also clear why *te*-lowering in this cluster type is infrequent, as well as why it applies mainly to V2. Its infrequency is due to the fact that for *te* to occur on V2, V1 has to move. Assuming that movement is more costly than no movement taking place, it is clear that *te*-in-situ should be preferred over *te*-lowering in this cluster type. The fact that *te* mainly lowers to V2 is also clear from the analysis presented above: it is the unvalued uT-feature on Mod which enters an Agree relation with the feature on F, and as a consequence can be head adjoined to F. V2 of the cluster, *blijven*, is situated in a lower Asp head, and its uT-feature can be directly valued by the [iT:Mod]-feature on Mod. This means that there is no feature on Asp that Agrees with a feature on F, which is the prerequisite for head movement to F and hence for *te*-lowering. As this never happens, *te*-lowering to V3 is also expected not to happen.

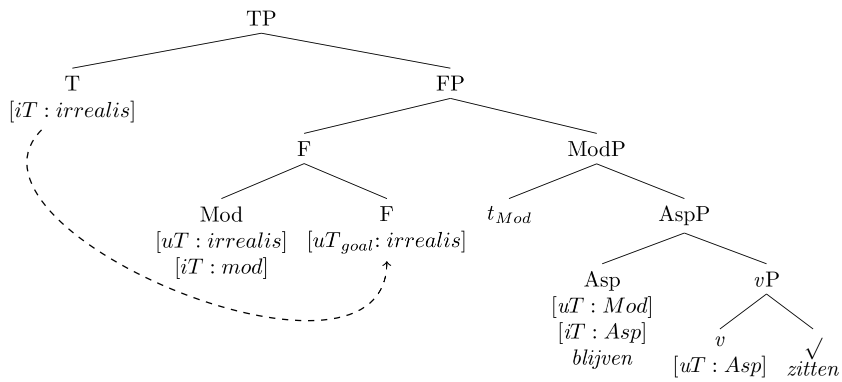
(118) **Cluster type Ia: lowering step I**



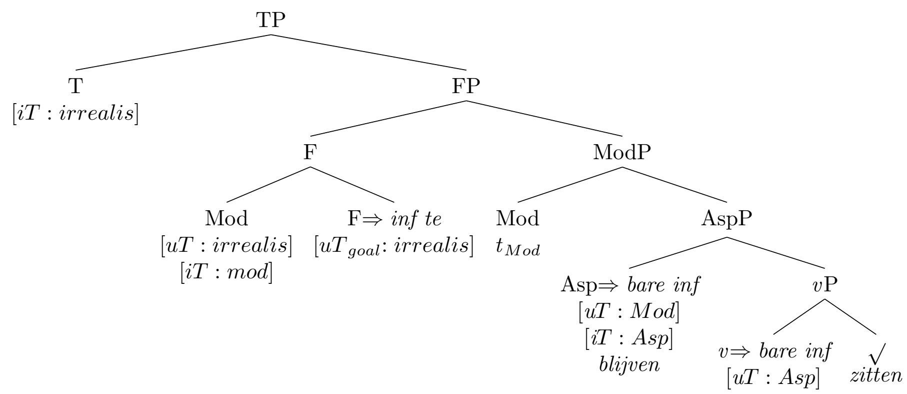
(119) **Cluster type Ia: lowering step II**



(120) **Cluster type Ia: lowering step III**

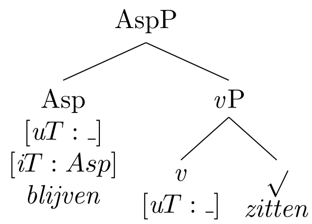


(121) **Cluster type Ia: lowering spell-out**



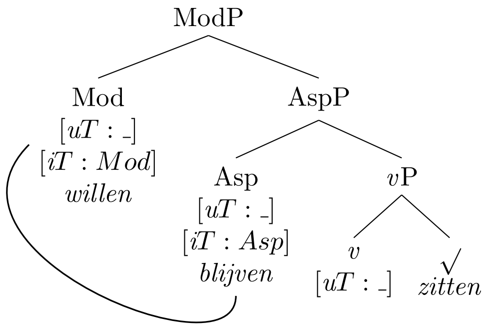
Let us finally look at *te*-doubling. The most frequent doubling pattern in this cluster is V1-*te*-V2-*te*-V3. I want to propose that cases of *te*-doubling in this cluster type result from the same verb movement of V1 in the cluster to head-adjoin to F, as we have seen above for *te*-lowering. On top of that, the *u*T-feature on *v* is also valued for *irrealis*, and can thus be spelled out as *te*. This results in the configuration V1-*te*-V2-*te*-V3. The question now is how *v* can get valued for *irrealis* in this cluster type, given that the first goal the *u*T-feature on *v* finds when probing up is the [*i*T:Asp]-feature on Asp. This first part of the derivation was given in (109) and is repeated here in (122).

(122) **Cluster type Ia: doubling step I**



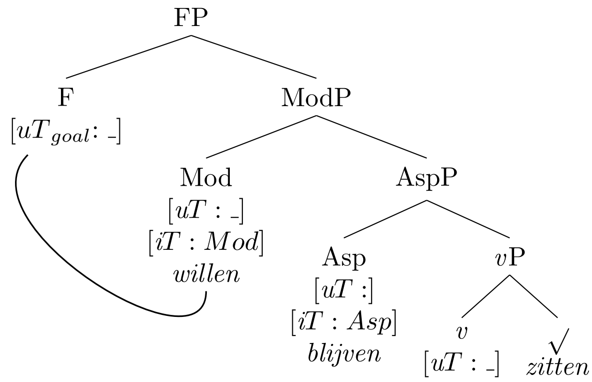
I want to propose that given that Asp bears two T-features, in some cases the *u*T-feature on *v* Agrees with the *u*T-feature on Asp, rather than the *i*T-feature on that head (see Müller (2009) for a similar explanation of variation caused by the order in which features are Agreed with). As I have argued above, the Agree relation between the two unvalued *u*T-features on *v* and Asp creates a connection between these two features, even though no valuation takes place. The same thing happens at the next step of the derivation when Mod is Merged: the *u*T-feature on Asp Agrees with the *u*T-feature on Mod rather than the *i*T-feature (123).

(123) **Cluster type Ia: doubling step II**

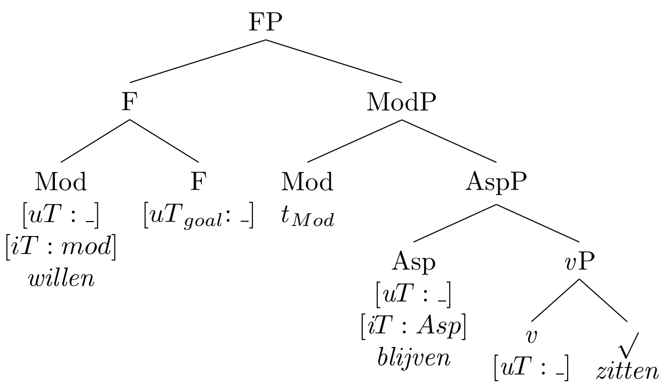


At the next step of the derivation, the uT-feature on Mod Agrees with the $u\mathrm{T}_{goal}$-feature on F (124). As a consequence, Mod Moves and head-adjoins to the left of F (125).

(124) **Cluster type Ia: doubling step III**

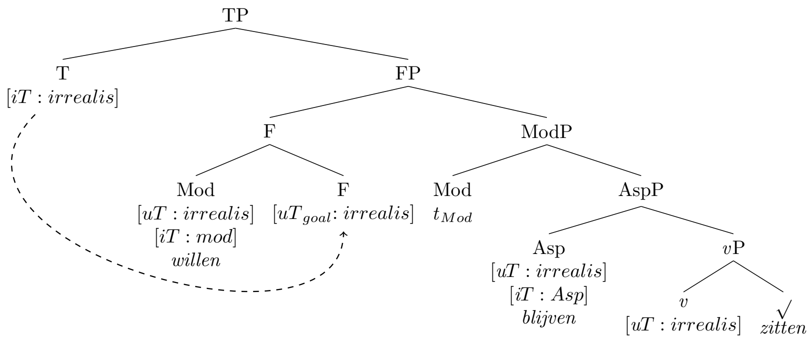


(125) **Cluster type Ia: doubling step IV**



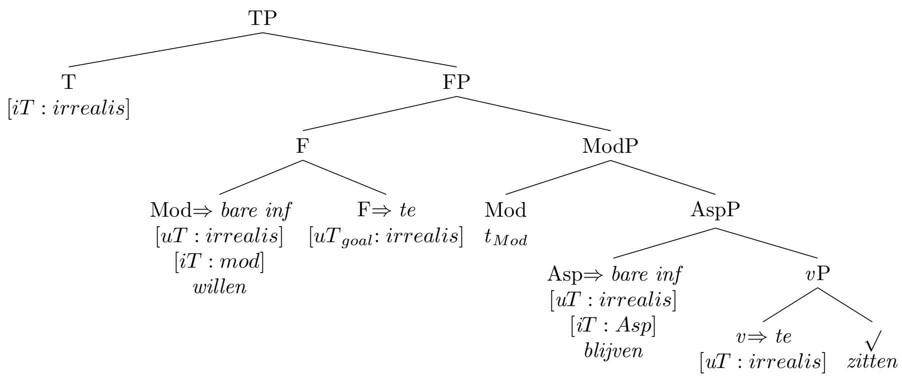
After that, T is Merged. Given that T only bears an iT-feature, the $u\mathrm{T}_{goal}$-feature on F will necessarily Agree with this iT-feature. As a consequence, it will be valued for *irrealis*, as will all the uT-features to which it is connected: the one on Mod, the one on Asp and the one on v (126).

(126) **Cluster type Ia: doubling step V**



Finally, recall that *te* can only be spelled out when it has its own syntactic head (F), or on v. This means that the spell out of the derivation given above results in the spell out of F as *te* and v as *te*, given that they both bear a [uT:irrealis]-feature. The other heads, Mod and Asp, even though they have the same feature, receive a default spell-out as a bare infinitive (127).

(127) **Cluster type Ia: doubling spell out**



The cluster thus has the form of the most frequent doubling configuration in this cluster: V1-*te*-V2-*te*-V3. Given that it involves both movement of a verb (V1) and Agree between a chain of uT-features, it is expected that this type of derivation is costly and therefore infrequent. This is indeed the case.

5.4.2 *Te*-placement and -presence in cluster type Ib

We now turn to cluster type Ib. The test sentence that was used in the questionnaire is repeated here in (101).

(128) Stijn **beweert** met het geld van zijn erfenis die grote villa
Stijn claims with the money of his inheritance that big villa
[*te* hebben$_1$ kunnen$_2$ kopen$_3$].
to have can buy
'Stijn claims he has been able to buy that big villa with his inheritance money.' CLUSTER TYPE IB

Recall from section 4.3.7 that in this cluster type (i) *te*-in-situ is accepted by all speakers, (ii) *te*-raising is excluded because *te* is already situated on the highest verb in the cluster (i.e. V$_1$ *hebben*), (iii)*te*-lowering is infrequent, and *te* mainly lowers to V$_2$ (almost 75% of the lowering cases) rather than to V$_3$, and (iv) *te*-doubling is also infrequent, with the most frequent configuration being V1-*te*-V2-*te*-V3 (almost 65% of the doubling cases). In addition, *te*-drop is regarded as noise in this cluster.

As I have done for the previous cluster, I first need to establish the status (functional, semi-lexically used or lexical) of the verbs in the cluster. V1 *hebben* 'have' is unquestionably a perfective auxiliary, and thus has a functional status. V3 *kopen* is the lexical verb, and therefore has a lexical status. V2 *kunnen* is also a functional verb, as can be seen in the following examples. It always occurs in IPP form when embedded under a perfective auxiliary (129), it rejects extraposition of the embedded infinitive (130), and it does not impose animacy restrictions on the subject (131).

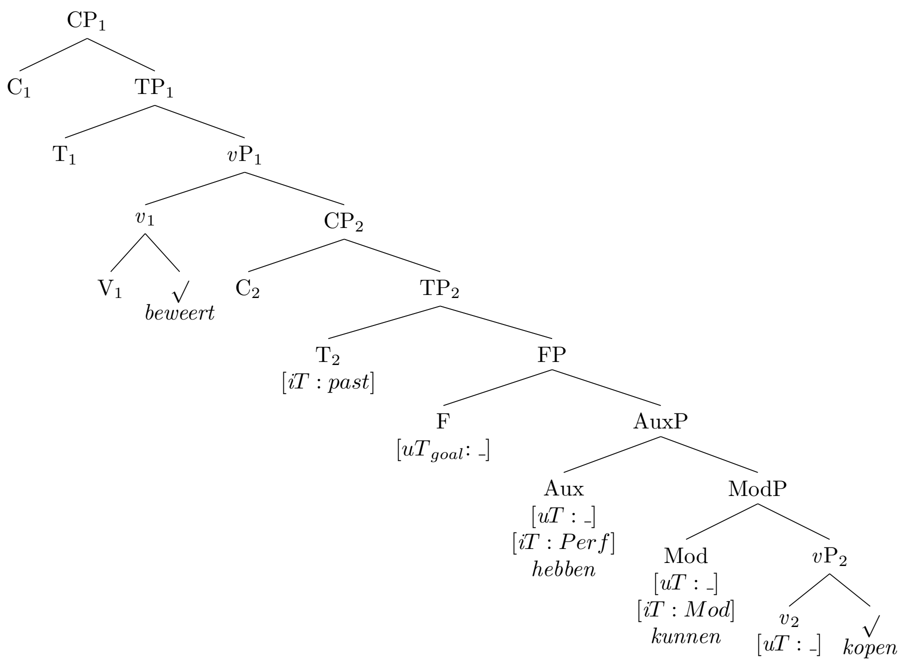
(129) ...dat hij een boek heeft **kunnen**/ ***gekund** kopen.
...that he a book has can.IPP/ can.PTCP go
'...that he managed to buy a book.'

(130) ...*dat hij heeft **kunnen**/ **gekund** [een boek kopen.]
...that he has can.IPP/ can.PTCP a book buy

(131) ***Het*** **kan** best sneeuwen morgen.
It can PRT snow tomorrow
'It is quite possible that it snows tomorrow.'

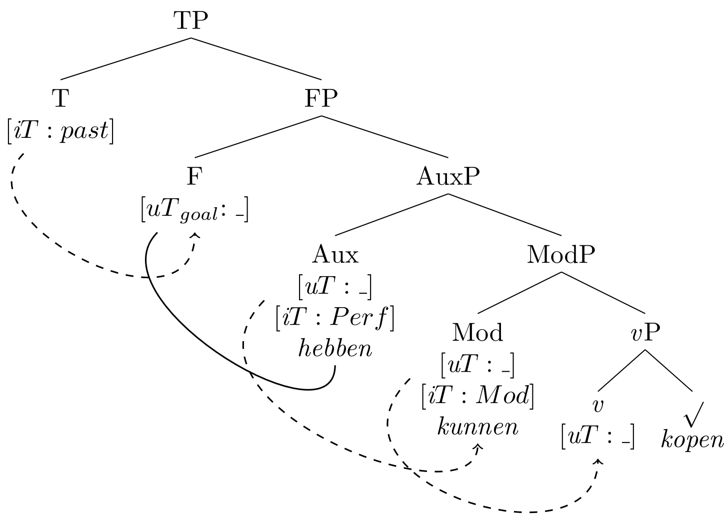
The verb that selects the *te*-infinitive in the test sentence is *beweren* 'claim'. This verb selects a complement that is at least as large as TP. In the test sentence in (128), the TP complement is has a reference to the past. The structure of the sentence is given in (132). All irrelevant projections (as well as any arguments) are left out for ease of exposition.

(132) **Cluster type Ib: structure**

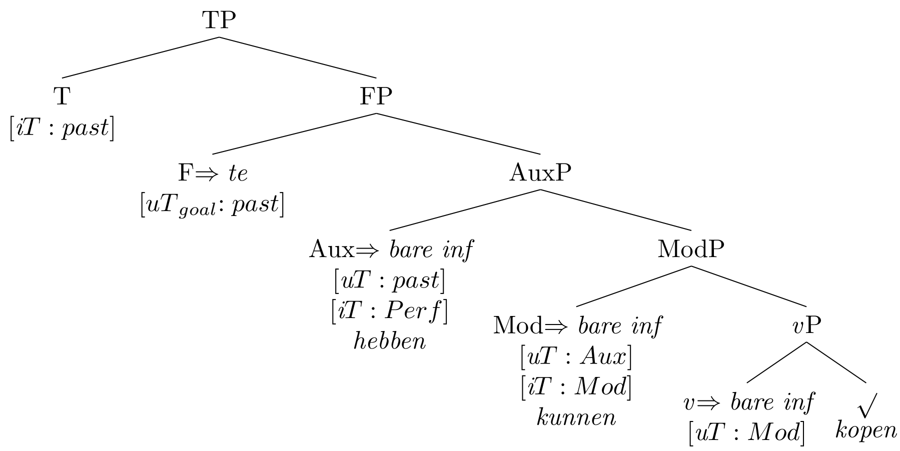


The feature valuation of the entire cluster is given in (133). All valuations are given in one structure, but note that Agree and the resulting valuation happens step by step, bottom up, as I have illustrated for cluster type Ia above. First, the unvalued *u*T on *v* is valued for Mod by the [*i*t:Mod]-feature on the next head. After that, the unvalued *u*T-feature on Mod is valued for Aux by the [*i*T:Aux]-feature on the next head. Then, the unvalued *u*T-feature on Aux probes up, but only finds the unvalued $u\mathrm{T}_{goal}$-feature on F. The Agree relation does not result in valuation, but establishes a connection between the two features. Lastly, the unvalued $u\mathrm{T}_{goal}$-feature probes up and finds the [*i*T:past]-feature on T. This feature values the $u\mathrm{T}_{goal}$-feature on F for *past*, and as a consequence, the connected *u*T-feature on Aux gets valued for *past* as well. The spell out of the cluster is given in (134). The F head is spelled out as *te*, due to the [$u\mathrm{T}_{goal}$:past]-feature on this head. Even though the *u*T-feature on Aux is valued for past, it will be spelled out as a default bare infinitive, as I have discussed above: the [*u*T:past]-feature can only be spelled out as *te* if it is hosted in a separate functional head F, or as verbal morphology in *v*. Similarly, even though the *u*T-feature on Mod is valued as Aux and should thus lead to Mod being spelled out as a past participle, the participial (verbal) morphology needs a *v* head to be spelled out on. This means that Mod gets a default spell out, namely as an infinitive. The *u*T-feature on *v* is valued for Mod, and thus gets spelled out as a bare infinitive. This leads to the 'correct' spell out of the cluster, with *te* on V1: *te-hebben*$_1$*-kunnen*$_2$*-kopen*$_3$. Recall that *te*-drop in this cluster is takenbeo eb noise, because its weighted frequency is below the threshold of 5%. As I have argued for cluster type Ia above, there is no reason why we would expect the Agree and valuation between [*i*T:past] on T and [*u*T$_{goal}$:_] on F to fail in this cluster, which means that we do not expect *te*-drop to be a frequent phenomenon.

(133) **Cluster type Ib: valuation**

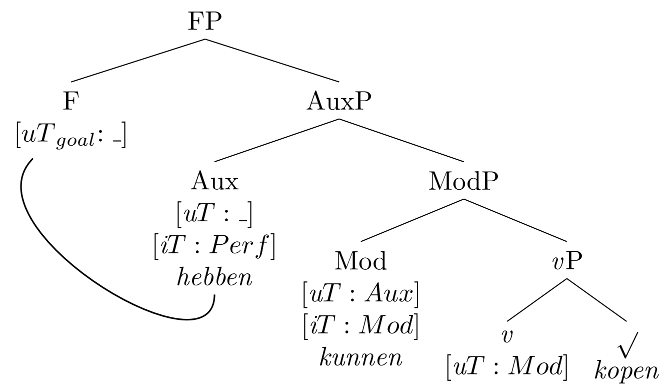


(134) **Cluster type Ib: spell out**

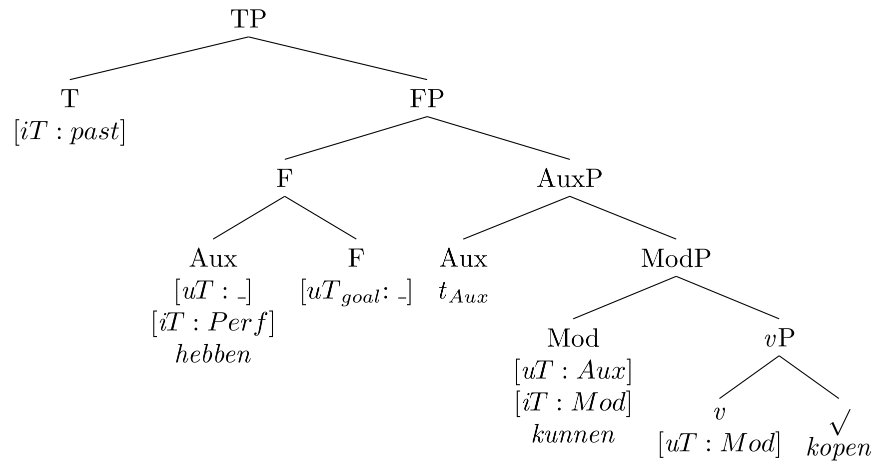


Te-lowering happens in the same way as I have argued for in cluster type Ia above. At the point in the derivation at which a link is established by Agree between the unvalued uT on Aux and the $u\mathrm{T}_{goal}$-feature on F (135), for some speakers this Agree relation results in movement of Aux to F, thus head-adjoining this head to the left of F (136). After that, as illustrated in (137), Agree between the $u\mathrm{T}_{goal}$-feature on F and the [iT:past]-feature on T results in the valuation of the former as *past*. Parasitically, the unvalued uT-feature on Aux gets valued for *past* as well.

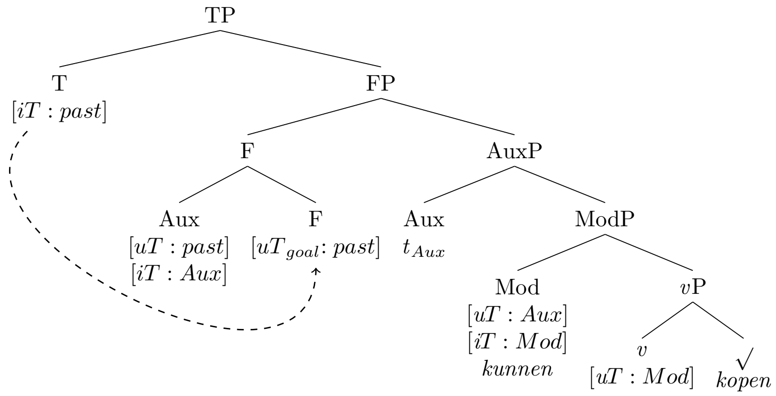
(135) **Cluster type Ib: lowering step I**



(136) **Cluster type Ib: lowering step II**

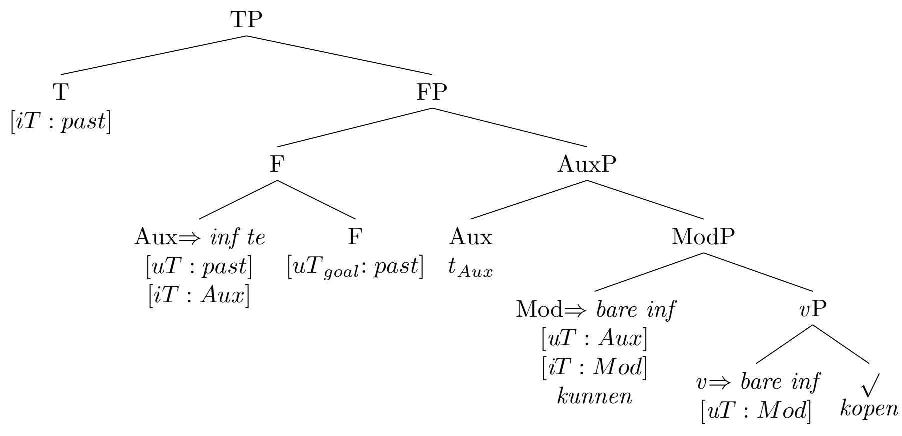


(137) **Cluster type Ib: lowering step III**



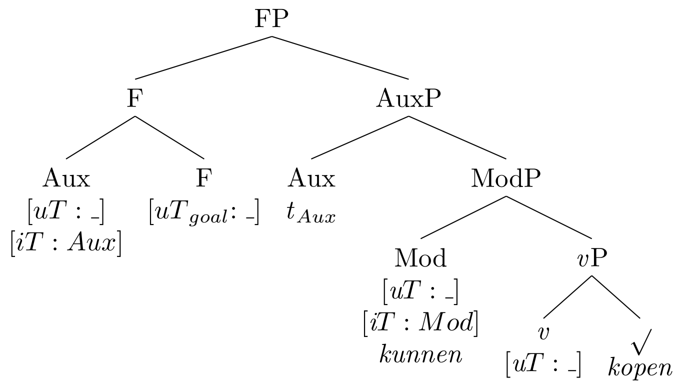
The spell-out is given in (138). As the [uT:past]-feature on Aux cannot be spelled out as *te*, Aux is spelled out as a bare infinitive. The F head is spelled out as *te*. Mod is spelled out as a bare infinitive, as there is no v head to host the participial morphology that would be spelling out [uT:Aux]. Lastly, v also gets spelled out as a bare infinitive, due to its [uT:Mod]-feature. This results in the lowered position of *te*: *hebben*$_1$*-te-kunnen*$_2$*-kopen*$_3$. As we have seen above for lowering in cluster type Ia, lowering in cluster type Ib mainly targets V2. This is what we expect based on the proposed derivation for *te*-lowering, because *te*-lowering is in fact the result of raising of V1 to F. This raising is triggered by an Agree relation between two unvalued uT-features on Aux and F. Given that no such Agree relation takes place between Mod and F, we do not expect *te*-lowering to V3 to take place.

(138) **Cluster type Ib: lowering spell out**

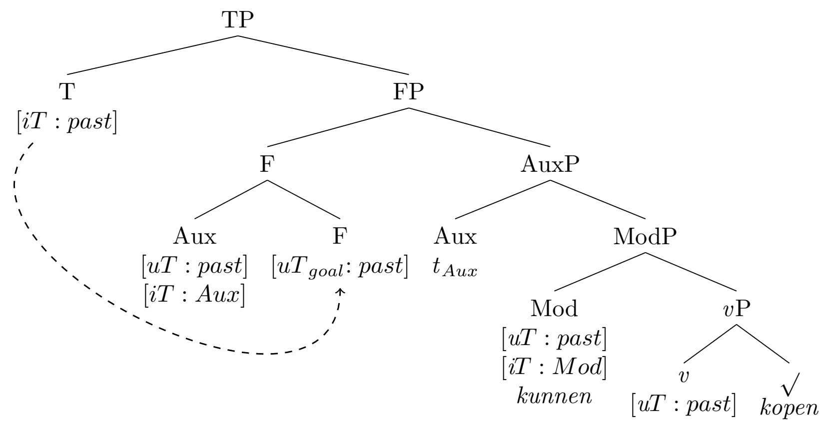


Let us finally look at *te*-doubling in this cluster. As was the case in cluster type Ia, the most frequent doubling configuration is V1-*te*-V2-*te*-V3. I therefore assume it results from the same derivation as I have proposed for *te*-doubling in cluster type Ia. That is, as with *te*-lowering, V1 (Aux) Moves and head-adjoins to F. Besides that, the *u*T-features on the lower heads Agree with each other rather than with the *i*T-features, resulting in a chain of connected *u*T-features (139). When the $u\mathrm{T}_{goal}$-feature on F finally Agrees with the *i*T-feature on T, and gets valued for *past*, all the connected *u*T-features are valued for *past* as well (140). Given that *te* needs either an F head or a *v* head as a host to be spelled out on, the result of this derivation is the cluster order V1-*te*-V2-*te*-V3 (141).

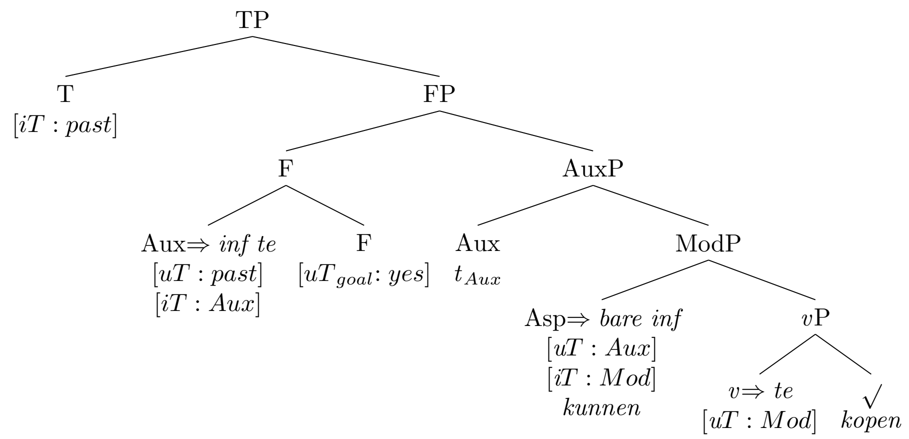
(139) **Cluster type Ib: doubling derivation**



(140) **Cluster type Ib: doubling valuation**



(141) **Cluster type Ib: doubling spell out**



5.4.3 *Te*-placement and -presence in cluster type II

We now move on to *te*-placement and -presence in cluster type II. The test sentence that was used in the questionnaire is repeated here in (142).

(142) Koen zal vanwege de winterstop vandaag niet [**hoeven**$_1$ *te*
Koen will because.of the winter.break today not need.INF to
gaan$_2$ voetballen$_3$].
go.INF play.football.INF
'Because of the winter break, Koen won't have to go and play football today.' CLUSTER TYPE II

Recall from section 4.3.7 that in this cluster: (i) *te* in situ is very frequent, (ii) *te*-drop is also frequent, (iii) as is *te*-raising, and (iv) *te*-doubling is infrequent, with the two most frequent configurations being *te*-V1-*te*-V2-V3 and *te*-V1-V2-*te*-V3. Based on the weighted frequencies, *te*-lowering is considered to be noise in this cluster. Recall furthermore that intraspeaker variation in this cluster is very high; many speakers allow two or three versions of this cluster, some even four (subsection 4.3.5).

Before I can give the full structure for this cluster, I have to establish the status of the verbs in the cluster. In subsection 5.2, I have shown that the *te*-selecting verb in this cluster, V1 *hoeven* 'need' is semi-lexical, and very close to being functional. Based on this fact, I assume that for some speakers *hoeven* is at the first stage of semi-lexicality, whereas it is at the second stage of semi-lexicality for others. This proposal is supported by (Van de Velde 2017:62-63), who shows that *hoeven* 'need' is increasingly selecting bare rather than *te*-complements, a transition which has started in the beginning of the 1950's. This means that *hoeven* is starting to have the same selectional requirements as the other functional modal verbs, which always select a bare infinitive. V3 *voetballen* is the lexical verb of the cluster, and V2 *gaan* has the status of a semi-lexically used verb. That is, on the one hand it always occurs in IPP form when embedded under a perfective auxiliary (143), and it rejects extraposition of the embedded infinitive (144). On the other hand, it imposes animacy restrictions on the subject. This is slightly complicated to show, as *gaan* is also used as a future auxiliary in Dutch. This future auxiliary is more grammaticalised than the motion verb *gaan*. The future auxiliary *gaan* can be used with a weather-*it* subject (145), but the motion verb *gaan* cannot (146). This means that the status of the motion verb *gaan* is semi-lexical.[18]

(143) ... dat hij een boek is **gaan**/ ***gegaan** kopen.
... that he a book is go.IPP/ go.PTCP buy
'... that he went to buy a book.'

(144) ... *dat hij is **gaan**/ **gegaan** [een boek kopen.]
... that he is go.IPP/ go.PTCP a book buy

(145) ***Het*** **gaat** morgen sneeuwen.
It goes snow tomorrow
'It will snow tomorrow.'

(146) ??***Het*** **gaat** naar beneden sneeuwen.
it goes to down snow
INTENDED: 'it goes and snows.'

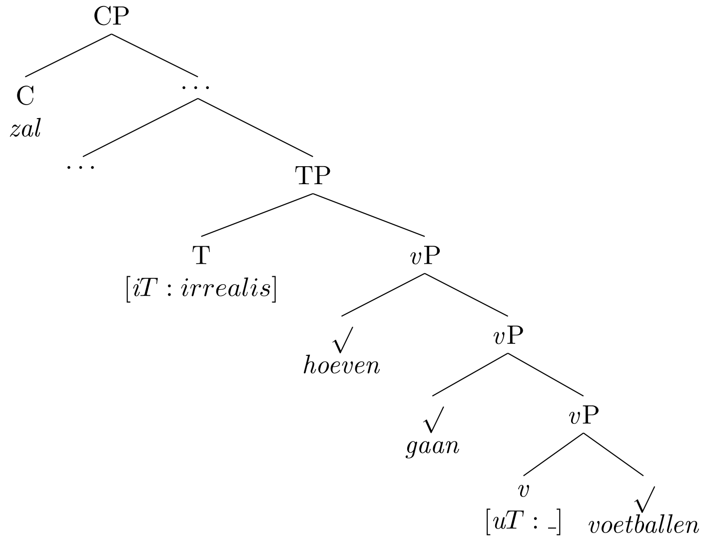
Given that I assume *hoeven* to be in a transitional stage between semi-lexically used and functional, I assume that across speakers *hoeven* is either still in the first stage of semi-lexicality, or already in the second stage. I will first discuss the analysis of the phenomena based on those speakers for whom *hoeven* is still in the first stage of semi-lexicality. After that, I will do the same for speakers whereby *hoeven* is already in the second stage of semi-lexicality.

The structure of the test sentence as used in the questionnaire is given in (147). All irrelevant projections are left out for ease of exposition. The structure of the sentence is a monoclausal one. I assume that the Dutch finite verb in main clauses is positioned in C (i.e. Verb Second position, following Bennis and Hoekstra (1989), Den Besten and Broekhuis (1989) and contra Zwart (1993) who argues that finite verbs in Verb Second position are sometimes in T and sometimes in C). The finite verb of the test sentence, *zal* 'will', is thus positioned in C. Given that the finite verb is a future auxiliary, the tense of the clause is *irrealis*; the playing of football has not been realised yet. The T of the clause thus bears an *i*T-feature which is valued for *irrealis*. V1 of the cluster is *hoeven*. In the first stage of semi-lexicality, the root of the verb is Merged in the

[18] Of course there is also the lexical verb *gaan*. When *gaan* is used lexically, its root is the most deeply embedded root of the verbal projection. When *gaan* is used semi-lexically, given the structures I have proposed in chapter 2, its root is part of the functional projection of another root, the latter being the most deeply embedded root of the verbal projection.

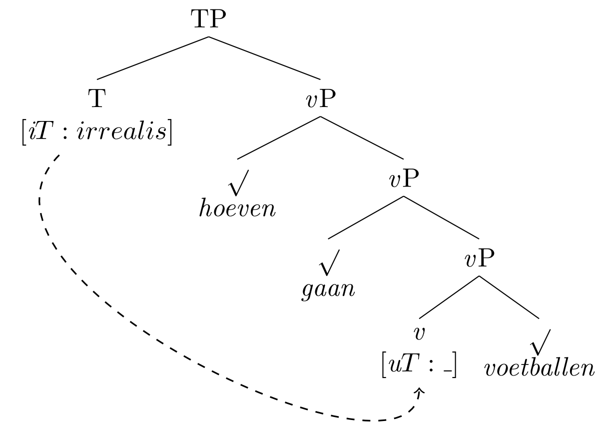
functional projection of the lower lexical verb. It cannot project, since it does not have any features. The same holds for V2 *gaan*, which also has the status of a semi-lexically used verb. The v of the clause bears an unvalued uT-feature. v projects, and does so twice more when each semi-lexically used root is Merged.

(147) **Cluster type II: structure (semi-lexical stage I)**



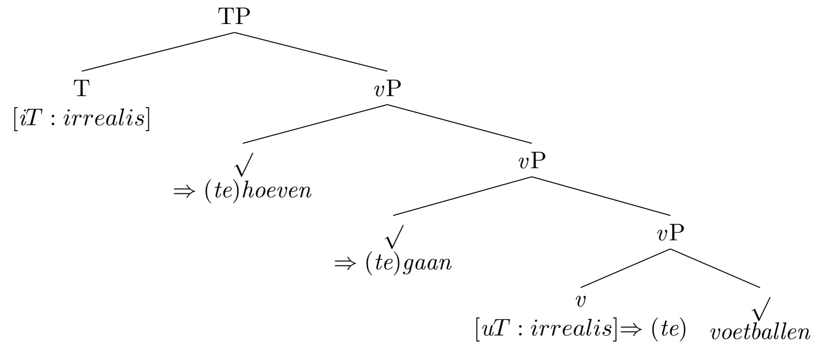
The feature valuation of the cluster is given in (148). When v is Merged, its unvalued uT-feature needs to become valued, i.e. v starts to probe up. However, the two next elements that are Merged are semi-lexically used verbs, V2 *gaan* and V1 *hoeven*, which are roots and thus do not bear any features. v continues to probe, and when T is Merged, the unvalued uT-feature on v can finally be valued by the [iT:irrealis]-feature on T. This means that the uT-feature on v is valued for *irrealis*. The spell out is given in (149). Based on (149), we would expect that *te* can be spelled out on either one of the verbs in the cluster, or even on multiple verbs. This is indeed more or less what we see reflected in the data. There is a large number of speakers that deals with the spell-out problem by spelling out the feature on the highest v projection. This results in the *te*-raising configuration: *te-hoeven*$_1$*-gaan*$_2$*-voetballen*$_3$. Another option is to spell the feature out on the middle projection, resulting in the *te*-in-situ configuration: *hoeven*$_1$*-te-gaan*$_2$*-voetballen*$_3$. A third option would be to spell out the feature on the lowest v projection, which would result in the *te*-lowering configuration: *hoeven*$_1$*-gaan*$_2$*-te-voetballen*$_3$. Recall from above, though, that this last option is below the threshold of 5% in frequency, and has thus been regarded as noise in the data. This means that spelling out the [uT:irrealis]-feature on the lowest v projection is the least preferred option of the three. Recall that the two doubling configurations that are the most frequent are *te*-V1-V2-*te*-V3 and *te*-V1-*te*-V2-V3. Speakers who accept *te*-doubling thus allow the [*u*T:irrealis]-feature to be spelled out on two of the *v* projections, but want one of those positions to be the highest of the three.[19] Another strategy for resolving the spell-out conflict in this structure could also be to not spell out the feature at all. That is, it might be that for some speakers, when the clause is sent to PF, the system rejects the spell out of the [*u*T:irrealis]-feature, as there is no straightforward position for this feature to be spelled out in. This would result in the *te*-drop configuration: *hoeven*$_1$-*gaan*$_2$-*voetballen*$_3$. Furthermore, I assume that the spell-out conflict in a structure like this leads to intraspeaker optionality. As it is sort of arbitrary on which projection *te* should be spelled out, it might very well be that at spell out, even one and the same speaker might not always choose the same option. The high intraspeaker variation regarding the placement and presence of *te* that was found for this cluster type is thus exactly what we would expect.

(148) **Cluster type II: valuation (semi-lexical stage I)**



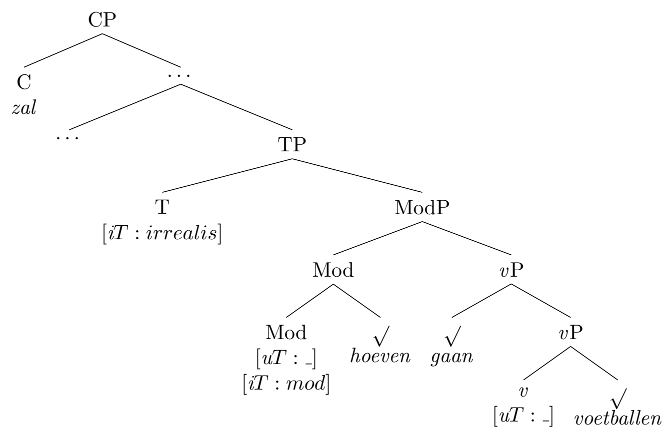
[19]Note that based on the structure in (149), we would expect *te*-tripling to be possible as well, namely by spelling out the *u*T:irrealis feature on all three projections of *v*. However, we have already seen that very few speakers spell out this feature in the lowest position, which makes it unlikely that *te*-tripling would ever be accepted. Furthermore, another factor which might rule out *te*-tripling is haplology: *te-hoeven-te-gaan-te-voetballen* might just be too cumbersome to pronounce.

(149) **Cluster type II: spell out (semi-lexical stage I)**



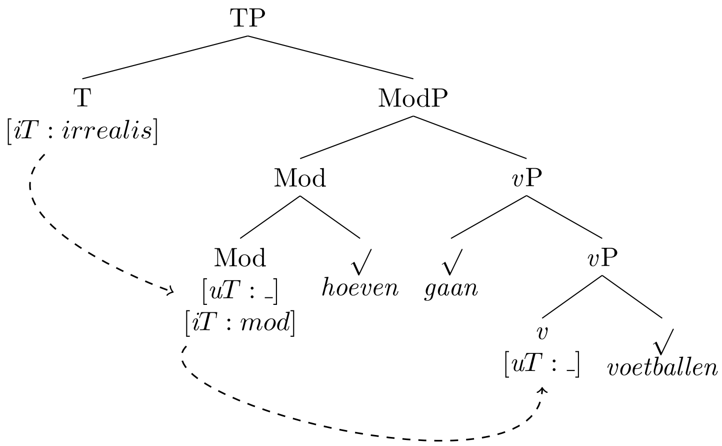
Let us now look at the structure, feature valuation and spell out of the cluster for those speakers in which *hoeven* is already in the second stage of semi-lexicality. At this stage, the semi-lexically used verb is Merged with a functional head in a separate workspace, and then inserted as a complex head in the functional projection of the lexical verb. In the case of *hoeven*, as already pointed out above, the root of *hoeven* is Merged with a Mod head. Aside from the complex Mod head, the structure of the sentence is the same as the one given above in (147). The structure with *hoeven* in the second stage of semi-lexicality is given in (150).

(150) **Cluster type II: structure (semi-lexical stage II)**



The valuation of the unvalued features is illustrated in (151). The first iT-feature that the unvalued uT-feature on v finds, is the [iT:Mod]-feature on Mod. As a result, it is valued as Mod. The unvalued uT-feature on Mod is valued for *irrealis* by the [iT:irrealis]-feature on T.

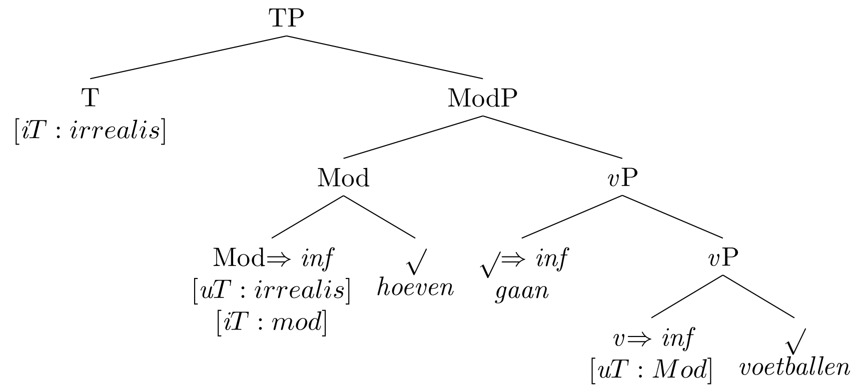
(151) **Cluster type II: valuation (semi-lexical stage II)**



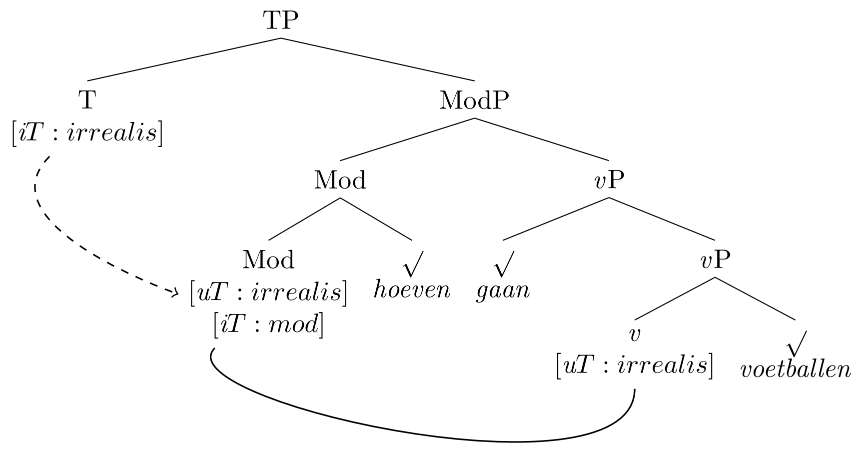
At spell out, as illustrated in (152), Mod, V1 *hoeven* receives a default spell-out as a bare infinitive, because the [uT:irrealis]-feature it bears needs an F head or v head to be spelled out on. Given that the uT-feature on v has been valued for Mod, and verbalises both the lower lexical root and the semi-lexically used root, V2 *gaan*, both roots are spelled out as bare infinitives. This results in the *te*-drop configuration: *hoeven*$_1$-*gaan*$_2$-*voetballen*$_3$. In other words, when *hoeven* is at the second stage of semi-lexicality for a given speaker, I expect this speaker to show *te*-drop in this cluster. Note, however, that there is one, more costly, way in which *te* could still be spelled out in this cluster. I have proposed above that for some speakers, the unvalued uT on v might optionally Agree with the unvalued uT-feature on a higher head, and will get valued only at a later stage in the derivation. In the case at hand, this would mean that the unvalued uT-feature on v would Agree with the unvalued uT-feature on Mod. When this latter feature is valued for *irrealis* by the higher [iT:irrealis]-feature on T, the uT-feature on v would be valued for *irrealis* as well.

The spell out of this alternative valuation would result in the possibility of *te* being spelled out on one (or both) of the v projections, or none, if the spell out conflict results in the [uT:irrealis]-feature being ignored at spell out. This is a more marked type of valuation, though, and I expect this to result in low frequencies of the spell out of such a derivation. This means that when *hoeven* is at the second stage of semi-lexicality, the unmarked option is the *te*-drop configuration. Only when a more marked valuation process has taken place will *te* be able to surface on one of the lower verbs in the cluster.

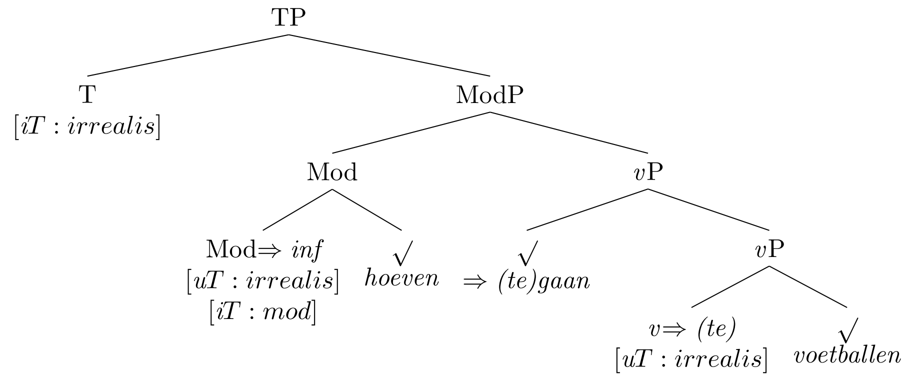
(152) **Cluster type II: spell out (semi-lexical stage II)**



(153) **Cluster type II: alternative valuation (semi-lexical stage II)**



(154) **Cluster type II: alternative spell out (semi-lexical stage II)**



5.4.4 *Te*-placement and -presence in cluster type III

Lastly, we turn to the placement and presence of *te* in cluster type III. The test sentence that was used in the questionnaire for this cluster is repeated here in (155).

(155) Peter zal vanwege de nieuwe dienstregeling binnenkort nog
Peter will because.of the new train.schedule soon even
langer op de trein [moeten$_1$ **zitten**$_2$ *te* wachten$_3$].
longer on the train must.INF sit.INF to wait.INF
'Because of the new train schedule, Peter will soon have to wait even longer for the train.' CLUSTER TYPE III

Recall from section 4.3.7 that in this cluster type (i) *te*-drop is by far the most frequent option, and is also obligatory for a large portion of the speakers, (ii) as a consequence of obligatory *te*-drop for many speakers in the sample, *te* in situ is much less frequent, (iii) *te*-raising is even less frequent, (iv) *te*-lowering is ruled out given that *te* is already on the lowest verb in the cluster (V3). Furthermore, based on a weighted frequency of less than 5%, *te*-doubling is considered to be noise for this cluster type.

Like I have done for the other cluster types, I first need to establish the status of the verbs making up the cluster. V1 *moeten* has a functional status, since it always appears in IPP form when embedded under a perfective auxiliary (156), rejects extraposition of the embedded infinitive (157), and does not impose animacy restrictions on the subject (158).

(156) ...dat hij een boek heeft **moeten**/ ***gemoeten** kopen.
...that he a book has must.IPP/ must.PTCP buy
'...that he had to buy a book.'

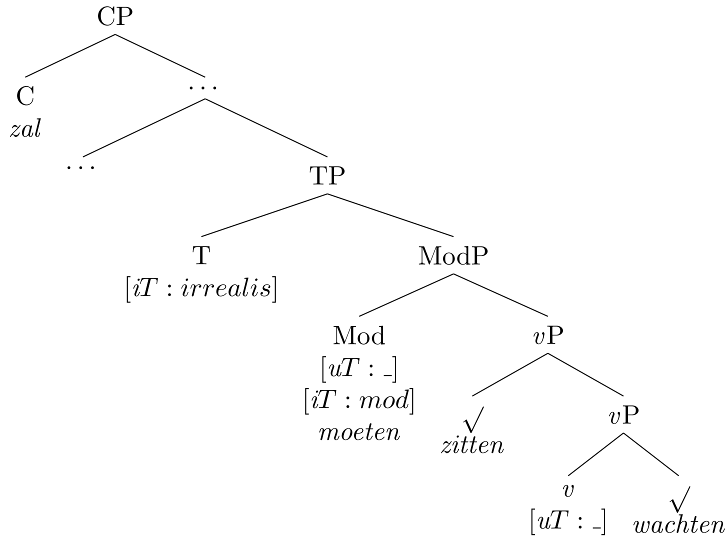
(157) . . . *dat hij heeft **moeten**/ **gemoeten** [een boek kopen.]
. . . that he has must.IPP/ must.PTCP a book buy

(158) ***Het*** **moet** deze winter echt veel sneeuwen.
It must this winter really a.lot snow
It really has to snow a lot this winter.'

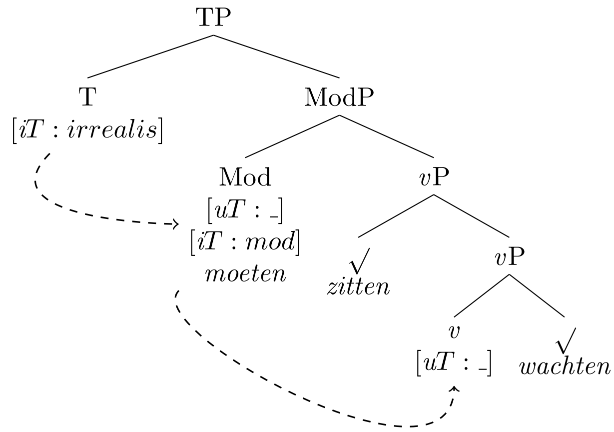
In subsection 5.2, I have determined that the *te*-selecting verb in this cluster, V2 *zitten*, is used semi-lexicallly. It does not show signs of becoming more functional, unlike what we saw for *hoeven*. For example, it still clearly imposes restrictions on the subject, and therefore rejects 'weather-*it*' subjects. I therefore assume that *zitten* is at the first stage of semi-lexicality. V3 *wachten* is the lexical verb of the cluster, and thus has a lexical status.

The structure of the test sentence that was used in the questionnaire is given (159). The structure is a monoclausal one. The finite verb *zal* 'will' is in C (Verb Second position). The tense of the sentence is irrealis: the 'having to wait' has not been realised yet. This means that the clausal T bears an [*i*T:irrealis]-feature. V1 of the cluster is a Mod head. Given that V2 *zitten* is at the first stage of semi-lexicality, it is a root that Merges with the *v*P of the verbalised lexical root. *v* bears an unvalued *u*T-feature, and projects twice. The valuation of the features in this cluster goes as follows. The *u*T-feature on *v* probes up, but cannot be valued by the semi-lexically used root of *zitten*, because it does not bear any features. When Mod is Merged, the *u*T-feature gets valued by the [*i*T:Mod]-feature on Mod. The *u*T-feature on this head gets valued for *irrealis* by the *i*T-feature on T. Both valuations are illustrated in (160).

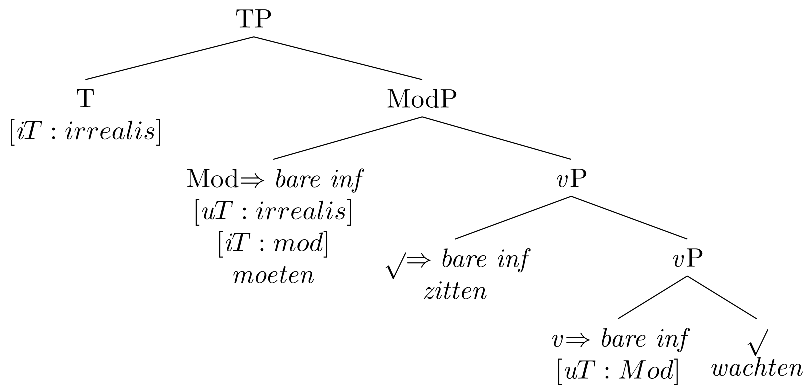
(159) **Cluster type III: structure (semi-lexical stage I)**



(160) **Cluster type III: valuation (semi-lexical stage I)**

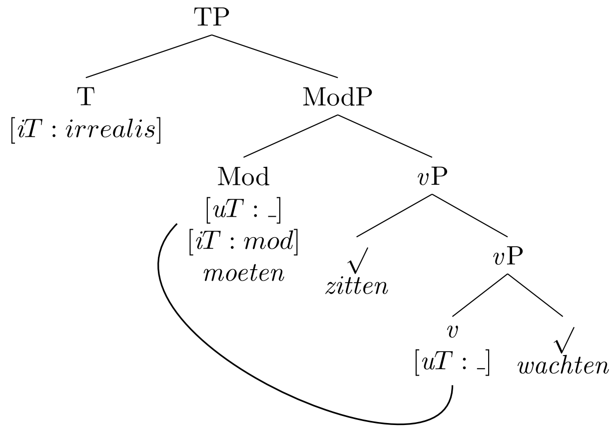


(161) **Cluster type III: spell out (semi-lexical stage I)**

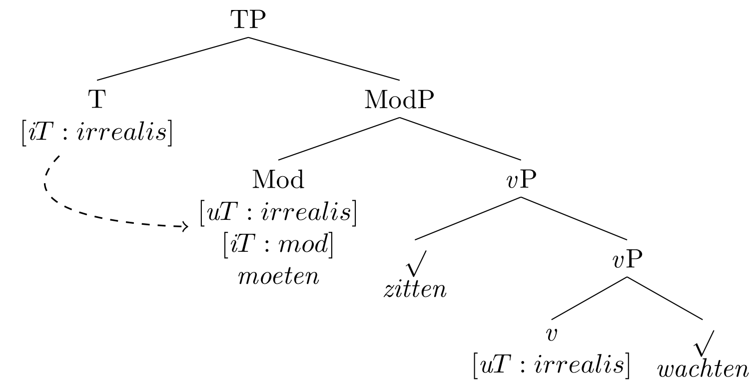


The spell out of the cluster is given in (161). As already discussed above, even though Mod bears a [*u*T:irrealis]-feature, it cannot be spelled out as a *te*-infinitive; *te* needs either an F head or *v* head to be spelled out on. Mod thus receives a default spell out, which means that it is spelled out as a bare infinitive. The [*u*T:Mod]-feature on *v* results in the spell out of the lexical verb (V3) as a bare infinitive as well. Because the projection with which the semi-lexically used root *zitten* (V2) was Merged is the same *v*, V2 *zitten* also gets spelled out as a bare infinitive. The result is the *te*-drop configuration: *moeten*$_1$-*zitten*$_2$-*wachten*$_3$. Recall from above that this is by far the most frequent option for this cluster type (a weighted frequency of more than 70%). This is thus exactly what we expect based on the structure and valuation of the features in this cluster. However, I still need to address the fact that, though much less frequent (weighted frequency 19,6%), *te* in situ occurs as well, as does *te*-raising (weighted frequency 7,7%). For these cases, I would like to propose a similar analysis as the one I have put forward for *te*-doubling in cluster types Ia and Ib. That is, I propose that for some speakers, the uT-feature on v Agrees with the uT-feature on Mod, rather than the [iT:Mod]-feature. This Agree relation creates a connection between the two unvalued uT-features. When the uT-feature on Mod gets valued by the [iT:irrealis]-feature on T, the uT-feature on v is also valued for *irrealis*. The two steps of this alternative valuation are given in (162) and (163).

(162) **Cluster type III: alternative valuation step one**

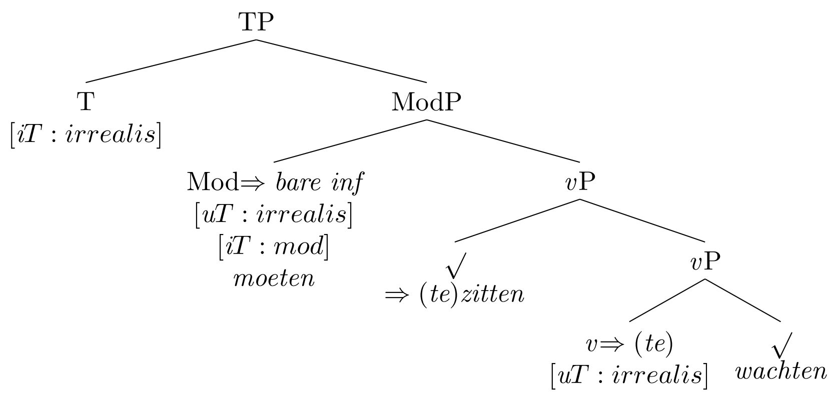


(163) **Cluster type III: alternative valuation step two**



The spell out of this alternative valuation is given in (164). Mod is spelled out as a bare infinitive (the default option). Like the first stage of semi-lexicality of *hoeven* in cluster type II, this alternative valuation results in a potential spell-out conflict. That is, the [*u*T:irrealis]-feature has two possible positions on which it can be spelled out, either on the root of the lexical verb, V3 *wachten*, or on the semi-lexically used root, V2 *zitten*.

(164) **Cluster type III: alternative spell out**



I thus expect that *te* can be either spelled out on V2, resulting in a *te*-raising configuration ($moeten_1$-te-$zitten_2$-$wachten_3$), or on V3, resulting in a *te*-in-situ configuration ($moeten_1$-$zitten_2$-te-$wachten_3$). Since this alternative valuation is marked, though, I do not expect these two phenomena to be very frequent in this cluster type, which is indeed the case. Note that there is also a third possibility for the spell out conflict, as I have argued for cluster type II above, namely that it results in the [*u*T:irrealis]-feature not being spelled out at all. In other words, one of the three possibilities for spell out after this alternative valuation is still a *te*-drop configuration. I thus expect that *te*-drop is by far the most common option in this cluster type, given that the standard valuation process never results in the spell out of *te* (cf. (161)), and one of the three options for spelling out the alternative valuation process also does not. Lastly, recall that *te*-doubling in this cluster type was taken to be noise, because its weighted frequency is below the 5% threshold. Technically, the alternative valuation could result in *te*-doubling, if the [*u*T:irrealis]-feature is spelled out at the level of both *v* projections. Apparently, this option is the least preferred way of dealing with the spell out conflict that results from the alternative valuation.

5.4.5 Explaining the groups of phenomena

In this last subsection of the analysis of the placement and presence of *te* in the four cluster types, I want to return to the results of the CA and HC, which suggested a particular grouping of phenomena. Before discussing how the present analysis can account for these groups, let me briefly recapitulate the findings (see 4.3.6.2–4.3.6.3). The output of the CA, and especially that of the HC (which is based on that of the CA), gave us three robust groups of phenomena.[20] These three groups of phenomena were:

1. *Te*-doubling in all cluster types
2. *Te*-in-situ in all cluster types, *te*-drop in cluster types II and III, and *te*-raising in cluster type II
3. *Te*-lowering in cluster types Ia and Ib

Let us therefore only look at the other three groups of phenomena. I first discuss the group of *te*-doubling in all cluster types. Note that based on the weighted frequency, I have taken *te*-doubling in cluster type III to be noise. *Te*-doubling in the other three cluster types is always infrequent and optional. In the analysis of the phenomena per cluster type, I have argued that *te*-doubling in cluster types Ia and Ib have exactly the same derivation and feature valuation. They both involve movement of V1, and an alternative valuation in which the uT-feature on v is valued for *irrealis* by T, as the result of a feature chain between the uT-features on v and the higher functional head. Because of this alternative valuation, v can be spelled out as *te* on V3. We have also seen that in cluster types Ia and Ib, there is a functional projection right above the highest verb of the cluster, which bears a $u\mathrm{T}_{goal}$-feature. This feature is valued for *irrealis* by the [iT:irrealis]-feature on T. In the case of *te*-doubling in these two cluster types, a [uT:irrealis]-feature is thus spelled out two times: once on F and once on v. Besides that, V1 has also moved and head-adjoined to the left of F. The result of this derivation is the most frequent *te*-doubling configuration in both cluster types: V1-*te*-V2-*te*-V3. *Te*-doubling in cluster type II is derived slightly differently. That is, the *te*-doubling configuration in this cluster type is the result of the [uT:irrealis]-feature on v being spelled out on more than verb in the cluster because the v bearing this feature projects multiple times. In other words, the doubling in cluster types Ia and Ib is the result of two separate [uT:irrealis]-features on separate heads being spelled out, while in cluster type II it is the result of one and the same [uT:irrealis]-feature being spelled out twice. Still, abstractly the spell out of these clusters boils down to the same thing, namely that [uT:irrealis] is spelled out twice. All cases of *te*-doubling are thus a robust phenomenon across cluster types in the sense that they are all the result of this feature being spelled out two times. The difference is that in cluster types Ia and Ib, an alternative valuation process has taken place. We would thus expect doubling in these two clusters to be slightly different from doubling in cluster type II. This is also the case if we look at the plot of the first two dimensions of the CA, here repeated in 5.1.

What we can see on this plot is that all doubling phenomena indeed group together. However, we can also see that doubling in cluster types Ia and Ib are closer to each other (and are thus more similar) than to doubling in cluster

[20] The output of the CA and HC gave a fourth 'group' of phenomena, existing only of *te*-drop in cluster type Ib. Recall that I have chosen to disregard the phenomena with a weighted frequency below 5%. The frequency for *te*-drop in cluster type Ib is below this threshold (4.3%). We can thus ignore this separate phenomenon as noise in the data.

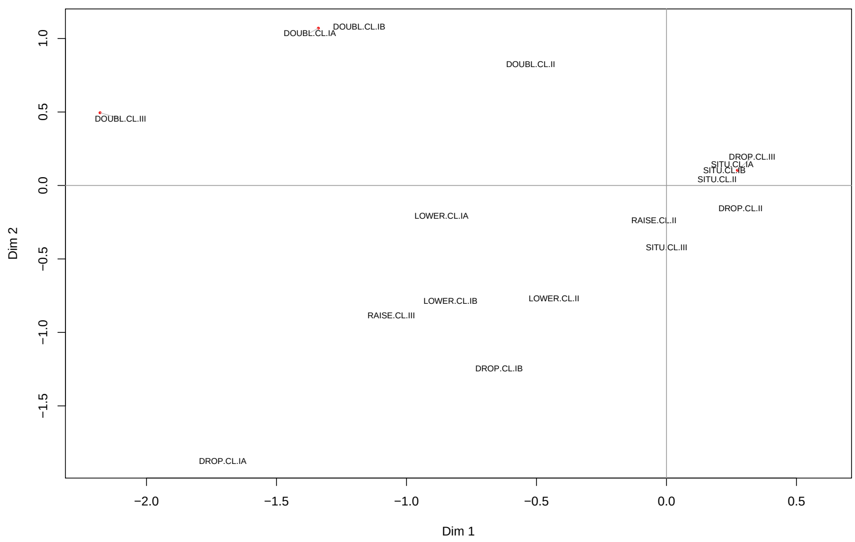


Figure 5.1: Dimensions 1 and 2 of CA (case study I)

type II. This more fine-grained distance between doubling in the former two compared to doubling in the latter is expected based on the analytical difference between them as outlined in the previous section (i.e. a standard valuation versus alternative valuation process).

Let us now move on to the next group of phenomena, which consists of *te*-in-situ in all cluster types, *te*-drop in cluster types II and III, and *te*-raising in cluster type II. In addition, *te*-raising in cluster type III partly belongs to this group. Based on the analysis I have presented above, we would definitely expect *te*-in-situ in cluster types Ia, Ib and II to cluster together with *te*-drop in cluster types II and III, and raising in cluster type II. We expect this because these are all the phenomena resulting from the most standard feature valuation processes in each cluster, without any additional movement having taken place. In cluster types Ia and Ib, this results in a *te*-in-situ configuration, because standard valuation will result in the valuation of uT on F as *irrealis*, this feature being spelled out as *te* right above the first verb of the cluster (V1). In cluster type II the situation is slightly more complex. When *hoeven* is at the first stage of semi-lexicality, uT on v is valued for *irrealis*. As v projects three times, over all three roots, *te*-raising *te*-in-situ, and *te*-lowering are all expected, the last one being the least preferred option (and thus not surprisingly having a frequency that is below the 5% threshold). Furthermore, *te*-drop is also an option, when the resolution of the spell out conflict results in the feature not being spelled out at all. When *hoeven* is at the second stage of semi-lexicality, *te*-drop occurs across the board when standard valuation has taken place. In other words, *te*-in-situ, *te*-raising and *te*-drop in cluster type II are all phenomena resulting from standard valuation in this cluster type. It is therefore expected that these phenomena form a group with the standard phenomena in cluster types Ia and Ib (i.e. *te*-in-situ). The same holds for *te*-drop in cluster type III: standard valuation in this cluster results in *te* not being spelled out. This group of phenomena can thus be seen as the phenomena for which no other valuation, movement, or double spell out has taken place. The odd one out is *te*-in-situ in cluster type III. As I have argued and illustrated in the previous subsection, *te*-in-situ in cluster type III requires an alternative valuation process, in which the uT-feature on v Agrees with the uT-feature on the higher Mod head, and then gets parasitically valued for *irrealis* because the uT-feature on Mod gets valued by the [iT:irrealis]-feature on T. Note, though, that if we once again look in a more fine-grained way at the plot of the first two dimensions of the CA, we can see that *te*-in-situ is the only in-situ phenomenon positioned below the x-axis, whereas all others are very close to each other near the origin. This is a subtle indication that *te*-in-situ in cluster type III is slightly different from *te*-in-situ in the other cluster types. A similar indication can be found if we return to the findings of the groups of speakers as reported in 4.3.6.3 above. Recall that I used HC applied to the participants of the questionnaire study rather than the phenomena. I investigated five groups of speakers, where I looked at the profile of so-called 'paragons'. These are the participants that are statistically the closest to the core of the group. Participant 1 turned out to have the following phenomena:

1. *te*-in-situ in cluster type Ia
2. *te*-in-situ in cluster type Ib
3. *te*-in-situ in cluster type II
4. *te*-raising in cluster type II
5. *te*-drop in cluster type III

These are all the phenomena that result from standard valuation, without movement or double spell out. As can be seen, *te*-in-situ in cluster type III is not a part of this list of phenomena. I take this to mean that this phenomenon does not completely belong to the group of 'standard' phenomena.[21]

Let us now look at the last group of phenomena: *te*-lowering in cluster types Ia and Ib. The fact that these two phenomena group together follows from the analysis, because both phenomena are the result of the same movement from V1 to F.

[21] Note that this paragon also does not have *te*-drop in cluster type II. This can be explained, though. For this speaker, *hoeven* must be in the first stage of semi-lexicality, because in the second stage we would expect him or her to only accept *te*-drop. Given that for this speaker *hoeven* is at the first stage, we expect him or her to accept both *te*-raising and *te*-in-situ. This is indeed what we find.

Besides the core groups of phenomena, there is one more finding that was discussed in subsection 4.3.6.2, which I want to address here. Based on the HC, it turned out that no matter how fine- or coarse-grained we do the clustering of the phenomena, we never end up with a cluster in which *te*-raising in cluster types II and III forms an exclusive group (i.e. without also including other phenomena). This finding can be explained now, because *te*-raising in cluster type II is the result of a standard valuation process, whereas for *te*-raising in cluster type III to be possible, an alternative, more marked valuation process needs to have taken place. In other words, the two *te*-raising phenomena have a different underlying valuation, which means that speakers who do not use the alternative type of valuation will only accept *te*-raising in cluster type II. Because their valuation is so different, it follows that they do not form a robust group of phenomena.

5.4.6 Different acquisition paths for *hoeven* and *zitten*

In the previous subsection I have analysed the presence and position of *te* in the four types of clusters that were tested in the questionnaire study. In the first two cluster types, Ia and Ib, the verb that selects the *te*-infinitive was the finite verb in Verb Second position, *zeggen* 'say' and *beweren* 'claim' respectively. Both verbs were shown to be non-restructuring verbs, selecting a complement of the size of at least a TP. In cluster type II, the verb that 'selects' the *te*-infinitive is *hoeven* 'need', and in cluster type III this is *zitten* 'sit'. Note that I put *select* in quotation marks here, as it turned out that these verbs, given that they are semi-lexically used and thus part of the functional domain of a lexical verb, do not actually select a *te*-infinitive any longer. They are verbs whose 'complement' has the size of a bare infinitive, leading to a monoclausal structure: *hoeven* and *zitten* are part of the functional structure of the lower lexical verb. We have thus seen that the two *te*-selecting verbs tested in this study really select for a TP/CP sized complement, whereas the other two are integrated in the functional domain of the lexical verb. Furthermore, we have also seen that there is a difference between the latter two verbs, *hoeven* and *zitten*. For *hoeven*, speakers show a high degree of optionality in the presence and placement of *te*. Furthermore, this verb is behaving more like a functional verb in the sense that it can be combined with weather-*it* subjects. In addition, Van de Velde (2017) shows based on a corpus study that since the 1950's there is a rapid increase in *hoeven* selecting a bare rather than a *te*-infinitive. By contrast, for *zitten* speakers show a much lower degree of optionality in the presence and placement of *te*. This verb also still imposes animacy restrictions on the subject. Based on these findings, I have analysed *hoeven* as a verb which is in an ongoing grammaticalisation process from the first stage of semi-lexicality to the second, whereas *zitten* is uniformly in the first stage.

In this subsection I present extra evidence for this analysis based on the relative frequencies of the lexical and semi-lexical use of *hoeven* and *zitten*. After presenting these frequencies, I propose that the way in which the semi-lexically use of *hoeven* and *zitten* is acquired can be different, using the Emergent Features model of Biberauer (2017a, 2019a).

I have done a corpus search in the Dutch corpus SoNaR+ (Oostdijk et al. 2013), in which I have searched for the semi-lexical use as well as the lexical use of *hoeven* and *zitten* respectively.[22] Let us start with *hoeven*. An example of its semi-lexical use is given here again in (165), and one of its lexical use in (166).

(165) Ik **hoef** niet naar huis **te gaan**.
I need not to home to go
'I don't need to go home.'

(166) Ik **hoef** geen cadeau.
I need no present
'I don't need a present.'

To return hits in which *hoeven* is used semi-lexically, I searched for (i) constructions in which *hoeven* is in Verb Second position, (ii) constructions in which *hoeven* is a finite verb in clause-final position, and (iii) constructions in which *hoeven* a non-finite verb in clause-final position. In all cases, *hoeven* takes an embedded infinitive. The clause final position can contain a single *te*-infinitive, or multiple infinitives. Examples are given in (167), (168) and (169) respectively.

(167) Ik **hoef** niet naar huis **te gaan** (fietsen).
I need not to home to go bike
'I don't need to go (and bike) home.'

(168) ... dat ik niet naar huis **hoef te gaan** (fietsen).
... that I not to home need to go bike
'... that I don't need to go (and bike) home.'

(169) ... dat ik niet naar huis heb **hoeven** (te) gaan (fietsen).
... dat I not to home have need.IPP to go bike
'... that I didn't need to go (and bike) home.'

To search for constructions like (167), I have used the query given in (170).

(170) $[lemma = \text{“}hoeven\text{”}][]\{1,5\}[word = \text{“}te\text{”}][pos\_head = \text{“}ww\text{”}]$

Searching for the *lemma* 'hoeven' returns all conjugations of this verb, be they in singular/plural, first, second or third person, simple present or simple past tense. The $[]\{1,5\}$ part makes sure there are one to five words in between *hoeven* in Verb Second position and the *te*-infinitive in verb final position. For example, in (167) there are three words in between *hoeven* and *te-gaan*. I have set the maximum at 5, to allow for more complex sentences than the one in

[22]The SoNaR+ corpus is a corpus of more than 500 million words. It consists of both written and spoken data, from the Netherlands and Flanders. For details, I refer the reader to *https://portal.clarin.inl.nl/opensonar_frontend/opensonar/about*.

(167), but not higher, to make sure there is not too much noise in the data. I have set the minimum at 1, and not 0, as the query that is presented next (in (171)) already returns strings in which there are no words intervening between *hoeven* and *te*. With [*pos_head* = "*ww*"] I make sure that the word to the right of *te* is a verb. To search for constructions like (168) and (169), I have used the query given in (171).

(171) [*lemma* = "*hoeven*"][*word* = "*te*"]?[*pos_head* = "*ww*"]

As with the previous query, searching for the lemma 'hoeven' makes sure that all conjugations of the verb are returned. The infinitive is included in this as well, meaning that this query also returns cases in which *hoeven* is embedded by a perfective auxiliary. By placing a question mark after the command [*word* = "*te*"], this word is indicated as being optional in the query. This means that cases of *te*-drop are also returned. To search for constructions in which *hoeven* is used lexically, I have used a very simple query, namely just the lemma 'hoeven' (172).

(172) [*lemma* = "*hoeven*"]

Because this query also returns all hits from the previous two queries (they contain [*lemma* = "*hoeven*"] as well), to get the total hits of the lexical use of *hoeven*, we need to subtract the total hits of the semi-lexical use of *hoeven* from the hits returned by the query in (172). The number of hits per query and their total frequency in the corpus are listed in Table 5.2.

	Total hits	Corpus percentage
Hoeven (Verb Second)	39.427	0.00716%
Hoeven (Clause final)	15.113	0.00275%
Hoeven (Total)	54.540	0.00991%
Hoeven (All hits)	81.933	0.0149%
Hoeven (lexical = 'all hits' - 'semi-lexical,total')	27.393	0.00499%

Table 5.2: Frequencies of semi-lexically used and lexical *hoeven*

As can be seen in the Table, is the total number of hits for the semi-lexical use of *hoeven* (54.540) almost twice as high as the total number of hits for the lexical use of *hoeven* (27.393) I hypothesise that the fact that the semi-lexical use of *hoeven* is so much more frequent than the one of its lexical use can play an enforcing role in the grammaticalisation process of *hoeven*. This hypothesis is supported by the findings from Van de Velde (2017), who shows that in the last seventy years *hoeven* has shown a drastic increase of selecting a bare rather than a *te*-infinitive.

Let us now turn to the frequencies of the semi-lexical and lexical use of *zitten*. An example of its semi-lexical use is given in (173), and of its lexical use in (174).

(173) Ik **zit** de hele dag **te werken**.
I sit the entire day to work
'I'm working the entire day.'

(174) Ik **zit** op de stoel.
I sit on the chair
'I'm sitting on the chair.'

As I have done for *hoeven* above, the semi-lexical use of *zitten* is searched for (i) as finite verb in Verb Second position, (ii) as finite verb in clause-final position, and (iii) as non-finite verb in clause-final position. In all cases, semi-lexically used *zitten* takes an embedded infinitive. The three configurations for semi-lexically used *zitten* are illustrated in (175), (176) and (177) respectively.

(175) Ik **zit** de hele dag **te werken**.
I sit the entire day to work
'I'm working the entire day.'

(176) ...dat ik de hele dag **zit te werken**.
...that I the entire day sit to work
'...that I'm working the entire day.

(177) ...dat ik de hele dag heb *zitten* (**te**) **werken**.
...that I the entire day have sit.IPP to work
'...that I've been working the entire day.'

To search for configurations like (175), I have used the following query (178).

(178) $[lemma = \text{“}zitten\text{”}][]\{1,5\}[word = \text{“}te\text{”}][pos\_head = \text{“}ww\text{”}]$

As was the case with *hoeven*, I searched for the lemma 'zitten' to return all conjugations of this verb. To allow for a number of words to appear between *zitten* and the clause final *te*-infinitive, I have used the same command again, allowing for a minimum of one and a maximum of five intervening words. To search for sentences like (176) and (177), I have used the query given in (179).

(179) $[lemma = \text{“}zitten\text{”}][word = \text{“}te\text{”}]?[pos\_head = \text{“}ww\text{”}]$

Once again, I have made the presence of *te* optional in the query. To search for the lexical use of *zitten*, I have used the following simple query (180).

(180) $[lemma = \text{“}zitten\text{”}]$

Because this query also returns all hits of semi-lexical *zitten*, I have calculated the total number of hits for the lexical use of *zitten* by subtracting the total hits for semi-lexically used *zitten*. The number of hits per query and their percentage in the corpus are given in Table 5.3.

From the table it becomes clear that the lexical use of *zitten* is much more

	Total hits	Corpus percentage
Zitten (semi-lexical, Verb Second)	25.509	0.00464%
Zitten (semi-lexical, clause final)	31.151	0.00566%
Zitten (semi-lexical, total)	56.660	0.0103%
Zitten (all hits)	539.541	0.09800%
Zitten (lexical = 'all hits' - 'semi-lexical,total')	482.881	0.0877%

Table 5.3: Frequencies of semi-lexically used en lexical *zitten*

frequent than its semi-lexical use, the former being close to five hundred thousand hits, whereas the latter is only around fifty five thousand hits. This is the reverse picture from the one we got with *hoeven* above. I hypothesise that the much lower frequency of the semi-lexical use of *zitten* than that of its lexical use is a factor that will not enforce rapid further grammaticalisation of semi-lexically used *zitten*. That is, the relative low frequency of its semi-lexical use will not 'help' this verb to grammaticalise from the first stage of semi-lexicality into the second stage of semi-lexicality.

Let us now move on to a proposal I want to make which implies a second difference between semi-lexically used *hoeven* and *zitten*. I want to propose that these two semi-lexically used verbs can be acquired differently. The acquisition path of *hoeven* can result in the language learner postulating that the semi-lexical use of this verb has the underlying structure of the second stage of semi-lexicality. The acquisition path of *zitten* never does. In order to develop my argument, I first discuss the Emergent Features model of Biberauer (2017a), on which the argument is based.

I follow Biberauer (2017a, 2019a) in assuming that syntactic features emerge in the grammar of a speaker through the process of language acquisition. In Minimalism it is standardly assumed that the set of possible syntactic features in natural language is regulated by UG. There is a universal set of features provided by UG, from which each language makes a 'one-time selection' in order to construct its lexicon (Chomsky 2001:10). Similarly, in Cartography (Cinque and Rizzi 2009), the set of possible features is provided by UG. Rather than making a selection from that set, however, languages differ in which of the UG-provided features are overtly expressed. So both in standard Minimalism and in Cartography, the set of possible features is provided by UG.

An even more Minimalist approach would be to assume that the features are not selected from a predetermined list in UG, but that they emerge from the process of language acquisition. This is the approach put forward by Biberauer (2017a, 2019a) (cf. also Zeijlstra (2008)). In this approach, UG only provides a general format for syntactic features, e.g. [iF] vs. [uF], or as attribute-value, [F: _], while the set of syntactic features in a language emerges during language acquisition from the interaction between this general format provided by UG, the

Primary Linguistic Data (PLD) and third factors. Biberauer (2017a) (building on Biberauer (2011) et seq) introduces a new model for language acquisition (and thus for variation and change as well), which takes into account the role and interaction of these three factors in the process of language acquisition. It can be schematically represented as in (181).

(181) *Three factors model of language acquisition*
UG (F_1) + PLD (F_2) + Maximise Minimal Means (MMM) (F_3) ⇒ I-language grammar (Biberauer 2016:1)

UG shapes language acquisition by providing the general format of interpretable versus uninterpretable features, i.e. [*i*F] vs. [*u*F]. As for the role of the PLD, Biberauer (2017a) makes a distinction between 'input', i.e. the language a child hears, and 'intake', which represents those aspects of the input the language learner is sensitive to, and which lead the learner to postulate syntactic features. The PLD has to be seen as 'intake', not 'input'. The aspects of the language input that the learner is sensitive to are the types of input that diverge from a direct one-to-one mapping between form and meaning (i.e. that diverge from Saussurian arbitrariness). These types of input signal to the learner that she cannot simply assume that the language system consists only of form/phonology and meaning/semantics, but that syntactic features should be postulated. Examples of such input include: (i) agreement, in which there are two forms, but only one meaning, (ii) structural Case, where there are forms without meaning, (iii) multifunctional morphemes, which signal that one form can have several meanings, (iv) ellipsis and other empty elements, in which there is meaning without form, and (v), movement, which can result in an extra, discourse-related or scopal, meaning. All these types of data in the PLD indicate to the language learner that their grammatical system cannot only consist of semantic and phonological features, and that syntactic [F]-features should be postulated to account for the systematic contrasts they find in their PLD. In other words, the syntactic features emerge through and during the process of language acquisition.

However, the emergence of features in the language of the language learner is not only regulated by the combination UG and the PLD, but also by third factors. The basic idea of Biberauer (2017a)'s third factors is that they constitute a general learning bias, which leads the learner to only postulate syntactic features when the PLD unambiguously points to the existence of these syntactic features. The overarching third factor principle is *Maximise Minimal Means* (MMM), of which there are two main manifestations: (i) Feature Economy (FE) as defined in (182), and (ii) Input Generalisation, given in (183).

(182) *Feature Economy (FE)*
Postulate as few syntactic features as possible to account for the input (=intake) (Biberauer 2017a:47)

(183) *Input Generalisation (IG)*
Maximise already-postulated features
(Biberauer 2017a:48)

FE operates on formal (syntactic) features, and essentially states that an learner should postulate as few syntactic features as possible in order to account for the intake, i.e. the (relevant part of the) PLD. This ensures that only those syntactic features are postulated that are unambiguously expressed in the PLD. In addition, the postulation of syntactic features 'piggy-backs' on the existence of semantic features (cf. Zeijlstra (2008)). IG on the other hand ensures that the features that are postulated are optimally put to use: it guides the learner to try to maximise the use of the already postulated features before positing a new one. Together, FE and IG result in an acquisition path in which the learner only postulates a new syntactic feature if the PLD provides clear evidence for this, and if the applicability of the existing features cannot be extended such that they include the new facts. The acquisition path thus proceeds in a NONE > ALL > SOME sequence, as illustrated in (184) (taken from Biberauer 2017a:48).

(184) **Acquisition path (Biberauer 2017a)**

F present?

NO (none) | YES: All heads?

YES (all) | NO: Which subset of heads?

↓

postulate a new [F] (some)

When the language learner receives input that does not unambiguously diverge from a one-to-one mapping between meaning and form, she will not postulate a syntactic feature; this is the 'none' option in path in (184). However, if the intake does signal such a divergence, she will postulate a syntactic feature [F], the interpretation of which is identical to the semantic feature involved in the form-meaning mismatch. IG now leads the learner to postulate the existence of this feature on *all* heads of the relevant type: the 'all' option in (184). If the learner then receives intake that clearly signals that not all heads of that category bear the previously postulated feature, she will postulate the existence of a new feature, present on a subset of the heads. This is the 'some' option in (184).

Summing up, in the approach of Biberauer (2017a), the set of syntactic features in a given language is not taken from a list of features provided by UG. Instead, the features are postulated by the learner based on unambiguous signals in the PLD (more specifically, mismatches between form and meaning), and thus emerge from the process of language acquisition. The set of syntactic features that is postulated is the smallest set that is compatible with the PLD, while also adhering to FE and IG. In a way, then, Biberauer (2017a)'s theory of Emergent Features is more minimalist than the standard Minimalist approach, in that it attributes less content to UG. Furthermore, their view is empirically more adequate, because it allows for more flexibility in the feature set of a language, thus leading to a more natural account of language variation and change. At the same time, the interaction between UG, the PLD and MMM also places severe limits on this variation and change. The theory is also immune to the Linking Problem (see among many others Pinker (1984), Gervain and Mehler (2010), Ambridge et al. (2013)), in that the question of how the learner maps the input from the PLD onto the syntactic features provided by UG is circumvented via the emergent nature of the syntactic features.

Returning now to the two semi-lexically used verbs under discussion here, *hoeven* and *zitten*, I want to show that the acquisition path can proceed in two ways for *hoeven*, whereas this is not the case for *zitten*. Let us start with *hoeven*. As I have shown in subsection 5.2, semi-lexically used *hoeven* behaves syntactically like the other Dutch modals in that it always appears in IPP form when embedded under a perfective auxiliary, rejects extraposition of the embedded infinitive, and does not have thematic restrictions. Where semi-lexically used *hoeven* differs from the class of modals, however, is the fact that it optionally embeds a *te*-infinitive, whereas modals can only embed a bare infinitve. The acquisition path for semi-lexically used *hoeven* can proceed in two ways. The first will result in the first stage of semi-lexicality. I assume that the learner will postulate a feature [Mod] present on the functional vocabulary items of all Dutch modals. For semi-lexically used *hoeven*, the learner will assume, based on the fact that this verb optionally select a *te*-complement, that it does not belong to the same category as the modals, but that it is the modal semantics of necessity of the root *hoeven* that brings about the modal interpretation when it is used semi-lexically. The fact that *hoeven* does not allow extraposition and always occurs in IPP form when embedded under a perfective auxiliary will indicate to the learner that *hoeven* is Merged in the functional domain of the lexical infinitive. Combining these two types of input will lead the learner not to postulate any formal feature on *hoeven*, but to assume that it is a root that is part of the functional structure of the lexical infinitive. This acquisition path will thus lead to semi-lexically used *hoeven* having the underlying syntactic structure of the first stage of semi-lexicality.

The second way in which the acquisition path of semi-lexically used *hoeven* can proceed is as follow. Recall that the only clear difference between semi-lexically used *hoeven* and the class of modals is the fact that the former optionally takes a *te*-infinitive. Recall also that the fact that it can select a *te*-infinitive has become less clear over the last decades, as shown by Van de Velde (2017). In other words, there is a group of Dutch speakers for whom semi-lexically used *hoeven* always selects a bare infinitive, like the modals. For these speakers, the difference between semi-lexically used *hoeven* and the other modals in terms of what type of complement they select is no longer sufficiently clear. The higher frequency of semi-lexically used *hoeven* taking a bare infinitive is part of the PLD of the new generation of language learners. Based on this intake, they will assume that semi-lexically used *hoeven* behaves syntactically very similar to the Dutch modals. This will lead language learners to analyse semi-lexically used *hoeven* as a root being Merged with a syntactic feature [Mod] in a separate workspace, after which it is Merged in the functional domain of the lexical infinitive. In other words: for these language learners semi-lexically used *hoeven* is at the second stage of semi-lexicality. Important to note here is that the class of modals are direct evidence for the existence of a [Mod]-feature in the language. This means that analysing the syntactic structure of semi-lexically used *hoeven* as forming a complex head with a [Mod]-feature satisfies both FE (no new feature is postulated) and IG (generalise over all modals).

In the case of *zitten*, the acquisition path only results in the first stage of semi-lexicality, due to several factors. First, in its semi-lexical use *zitten* does not belong to an identifiable subset of verbs with similar syntactic behaviour and related semantics, unlike what is the case for *hoeven* and the Dutch modals. In other words, there is no subset of verbs in Dutch that will lead the learner to postulate a formal feature like [$\text{Asp}_{progressive}$] or [$\text{Asp}_{durative}$]. The other verbs that, like *zitten*, can bring about this type of interpretation, i.e. the other two posture verbs *liggen* 'lie' and *staan* 'stand', and the motion verb *lopen* 'walk', are all in an early semi-lexical stage, in the sense that they thus reject weather-*it* subjects (185) (see also subsection 2.3.3).

(185) a. ****Het*** **loopt** te sneeuwen.
it walks to snow
b. ****Het*** **staat** te sneeuwen.
it stands to snow
c. ****Het*** **ligt** te sneeuwen.
it lies to snow

There are two other ways in Dutch to create a progressive interpretation, the BE AT THE INF construction and the BE BUSY WITH construction, illustrated in (186) and (187) respectively.

(186) Ik **ben aan het** schrijven.
I am at the writing.INF
'I'm writing.'

(187) Ik **ben bezig met** schrijven.
I am busy with writing
'I'm (busy) writing.'

In these constructions, however, there is no single functional vocabulary item or underlying syntactic head that can be assumed to bear an [$\text{Asp}_{progressive}$]-feature; it is the combination of all elements of the construction that brings about a progressive interpretation. This means that the language learner does not have any evidence for postulating the existence of an [Asp$_{progressive}$]-feature in the language. In other words, in the case of *zitten* (and also in the case of *liggen*, *staan* and *lopen*), the language learner will not postulate that these roots can be Merged with a formal feature to bring about the progressive interpretation of the lexical verb they combine with. The fact that *zitten* always appears in IPP form when embedded under a perfective auxiliary and rejects extraposition of the embedded infinitive will lead the learner to assume *zitten* is part of the functional domain of the lexical infinitive. That is, he or she will assume the structure of the first stage of semi-lexicality for *zitten*.

Summing up, the different relative frequencies of the lexical use of *hoeven* versus *zitten*, as well as the different featural acquisition path for both verbs, indicate that *hoeven* for some speakers/learners has reached the second stage of semi-lexicality, whereas for *zitten* all evidence points towards the first stage.[23] These findings support the analysis as presented in subsections 5.4.3–5.4.4, in which the high degree of optionality in the presence and placement of *te* with *hoeven* is the result of this verb being in a transition phase from the first stage of semi-lexicality to the second, whereas *zitten* is uniformly in the first stage.

[23] A fruitful direction for future research here would be to test my proposal using acquisition data. This could be done by a corpus study in the CHILDES corpus, which contains large sets of L1 acquisition data, as well as child directed speech by parents.

CHAPTER 6

Conclusion and outlook

In this case study I have presented a study into the effects of the degree of semi-lexicality of two verbs, *hoeven* 'need' and *zitten* 'sit' on the presence and placement of *te* in Dutch non-finite three-verb clusters. Using a large-scale questionnaire study, I have investigated variation and optionality regarding these phenomena in four different cluster types. Before discussing some implications of the analysis of this case study, let me briefly recapitulate the main findings. Four non-standard or partly non-standard phenomena were discovered: *te*-raising (i.e. *te* appearing on a verb higher than the one required by selection), *te*-lowering (i.e. *te* appearing on a verb lower than the one required by selection), *te*-drop (i.e. *te* being absent despite it being required by selection), and *te*-doubling (*te* occurring twice in the cluster, whereas selection requires only one *te*). For *te*-raising and *te*-drop, I have found that the frequencies vary greatly between different cluster types. Furthermore, these two phenomena, but especially *te*-drop, can sometimes even be obligatory for a speaker. *Te*-lowering and *te*-doubling both occur quite infrequently, but more or less with the same frequencies across cluster types.

I have argued that we have to distinguish between *te*-selecting verbs that select a large *te*-complement of at least the size of a TP, and *te*-selecting verbs that take a complement that is as small as a bare infinitive. In the first case, *te* is a functional head in the syntactic structure, which is biclausal. The *te*-selecting verbs of cluster types Ia and Ib, *zeggen* 'say' and *beweren* 'claim' respectively, fall in this category of verbs that select at least a TP complement. In the second case, the structure of the *te*-selecting verb and the embedded infinitive is monoclausal, and *te* is not a functional head in the structure, but the spell-out of a uT-feature on v, when this feature has been valued for *irrealis* or *past*.

The '*te*-selecting' verbs of cluster types II and III, *hoeven* 'need' and *zitten* 'sit' respectively, fall in this category of verbs that 'select' a complement the size of a bare infinitive. As pointed out in the analysis, these verbs strictly speaking no longer c-select a *te*-complement. *Hoeven* and *zitten* are semi-lexically used verbs that are Merged into the functional domain above the lexical infinitive. The underlying syntactic structure of the clusters of *hoeven* in semi-lexical stage I results in a spell out conflict for a [*u*T:irrealis]-feature to be spelled out, leading to the absence or unexpected placement of *te*. When this verb is at the second stage of semi-lexicality, the underlying syntactic structure of the cluster almost always results in *te*-drop. This also holds for the underlying structure of the cluster of semi-lexically used *zitten*. In the rest of this chapter, I want to discuss three avenues for future research that follow from this case study.

The first direction for future research relates to *te*-presence and placement with three other Dutch restructuring verbs, namely *durven* 'dare', *beginnen* 'begin' and *proberen* 'try'. In subsection 5.4.2, I have pointed out that *hoeven* and *zitten* are not the only verbs that allow *te*-drop. That is, many speakers allow *te*-drop with *durven*, *beginnen* and *proberen* as well. As far as I have been able to ascertain, this holds for both the Netherlands and Flanders for *durven*, and only for Flanders for *beginnen* and *proberen*. One constraining factor on *te*-drop with these verbs is that it is only possible in real clustering contexts, i.e. Verb Raising constructions, and not in the Third Construction or Extraposition construction. The relevant examples are repeated in (1)–(3).

(1) a. ...dat ik *het boek* niet heb **durven** **(*te*)** lezen. VR
...that I the book not have dare.IPP to read
b. ...dat ik *het boek* niet heb **gedurfd** ***(*te*)** lezen. 3C
...that I the book not have dare.PTCP to read
c. ...dat ik niet heb **gedurfd** (**om**) *het boek* ***(*te*)** lezen. EP
...that I not have dare.PTCP for the book to read
'...that I haven't dared to read the book.'

(2) a. ...%dat ik *het boek* niet ben **beginnen** **(*te*)** lezen. VR
...that I the book not am begin.IPP to read
b. ...dat ik *het boek* niet ben **begonnen** ***(*te*)** lezen. 3C
...that I the book not am begin.PTCP to read
c. ...dat ik niet ben **begonnen** *het boek* ***(*te*)** lezen. EP
...that I not have begin.PTCP the book to read
'...that I haven't begun to read the book.'

(3) a. ...%dat ik **het boek** niet heb *proberen* **(*te*)** lezen. VR
...that I the book not have try.IPP to read
b. ...dat ik **het boek** niet heb *geprobeerd* ***(*te*)** lezen. 3C
...that I the book not have try.PTCP to read
c. ...dat ik niet heb *geprobeerd* (**om**) **het boek** ***(*te*)** lezen. EP
...that I not have try.PTCP for the book to read
'...that I haven't tried to read the book.'

In 5.4, I have analysed *te*-drop in a monoclausal structure as the result of a spell-out conflict: [*u*T:irrealis] on *v* can be spelled out on several projections of this *v*, and for some speakers this conflict is resolved by not spelling the feature out at all. If *durven*, *beginnen* and *proberen* are part of a monoclausal structure in a Verb Raising construction, my analysis would predict that in these constructions, some speakers should also allow *te* to be raised. Examples are given in (4)–(6).[1]

(4) ?Hij zal morgen wel weer ***te* durven** gaan voetballen.
He will tomorrow PRT again to dare.INF go.INF play.football.INF
'Tomorrow he will probably dare to go and play football again.'

(5) ??Hij zal morgen wel weer willen ***te* beginnen** voetballen.
He will tomorrow PRT again want.INF to go.INF play.football.INF
'Tomorrow he will probably want to start again with playing football.'

(6) ?Hij zal morgen wel weer ***te* proberen** gaan voetballen.
He will tomorrow PRT again to dare.INF go.INF play.football.INF
'Tomorrow he will probably dare to go and play football again.'

Unfortunately, these types of sentences were not tested in the questionnaire study. In my opinion, *te*-raising with *durven* and *proberen* sounds more acceptable than with *beginnen*, but that might be due to the fact that with *beginnen* I prefer the extraposition construction over the Verb Raising construction to begin with. A future questionnaire study is needed to test whether *te*-raising is possible with these verbs when they are embedded as an infinitive in a Verb Raising construction, and furthermore whether *te*-raising is accepted in the same locations as *te*-drop with these verbs. It should also test the acceptance of the three possible constructions, Verb Raising, the Third Construction and Extraposition, with these three verbs among the speakers, to investigate whether a higher preference for for example the Extraposition construction would block *te*-drop or *te*-raising with the same verb in Verb Raising contexts. That is, if in a specific part of the Dutch language area Extraposition is the most frequent option for one of these verbs, the language learner will not get enough intake that points to these verbs also being able to occur in a monoclausal construction, and s/he will therefore not assume an optional underlying structure in which *te*-drop or *te*-raising is possible.

The second direction for future research relates to a difference in the optionality of *te* when 'selected' by *hoeven* 'need' and *zitten* 'sit' in Verb Second position compared to a position in the verb cluster. Throughout the case study, we have seen that for many Dutch speakers, *te* can or needs to be absent when *hoeven* and *zitten* are an infinitival verb in a verb cluster at the end of the clause, as illustrated again in (7) and (8).

[1] In (5), *beginnen* is the second verb of the three-verb cluster, because it cannot embed lower aspectual verbs, like *gaan* 'go' in (4) and (6).

(7) Koen zal vandaag niet [**hoeven** **(*te*)** gaan voetballen].
Koen will today not need.INF to go.INF play.football.INF
'Koen won't have to go and play football today.'

(8) Peter zal nog langer op de trein [moeten **zitten** **(*te*)** wachten].
Peter will even longer on the train must.INF sit.INF to wait.INF
' Peter will have to wait even longer for the train.'

However, when semi-lexically used *hoeven* and *zitten* are the finite verb in Verb Second position, *te* becomes obligatory (among others Haeseryn et al. (1997), IJbema (2001), Van Pottelberge (2002), Dreumel and Coppen (2003)). This is illustrated in (9) and (10).

(9) Koen **hoeft** vandaag niet [***(*te*)** gaan voetballen].
Koen needs today not to go play.football
'Koen doesn't need to go play football today.'

(10) Peter **zit** lang op de trein [***(*te*)** wachten].
Peter sits long on the train to wait
'Peter is waiting for the train for already a long time.'

At the moment, I do not have a straightforward analysis for these facts. However, I do want to mention here that this strong requirement of *te*-presence when *hoeven* and *zitten* occur as the finite verb is weakened when these verbs are used as in a position where they are positionally very close to the lexical infinitive (cf. also Haeseryn et al. (1997), Dreumel and Coppen (2003)), and furthermore morphologically resemble a bare infinitive. For example, compare (10) and (9) with (11) and (12).

(11) ...dat ze vandaag niet [**hoeven** **(*te*)** gaan voetballen.
...that they today not need to go play.football
'...that they won't have to go play football today.'

(12) ...dat ze vandaag lang op de trein [**zitten** **(*te*)** wachten.]
...that they today long on the train sit to wait
'...that they're waiting for the train for a very long time today.'

In these sentences, the finite plural form of *hoeven* and *zitten* looks on the surface exactly the same as their infinitival forms. Furthermore, the verbs are not separated by any other element than *te* from the lexical infinitive. It thus seems that when *hoeven* and *zitten* have been Moved to a much higher position (i.e. Verb Second position) in the syntactic structure than their Merge position right above the lexical infinitive, the syntactic structure does not lead to a spell out conflict for *te* to be spelled out, whereas when these verbs are close to the lexical infinitive, it does. In other words, *te* always appears when *hoeven* and *zitten* are syntactically far away from the lexical infinitive. A detailed questionnaire study related to the presence of *te* in Verb Second versus clause-final position is needed to get a full understanding of these data.

A third direction for future research is investigating different cluster orders in non-finite three-verb clusters, and the effect the different word orders has on the presence and placement of *te*. Recall from the introduction chapter of this case study (chapter 3), that in Dutch the order of verbs in finite three-verb clusters is well-documented and analysed. The possibilities relating word order in non-finite three-verb clusters, however, has hardly been investigated, nor the position and placement of *te* in these latter types of clusters. In subsection 4.3.7, I already mentioned and presented part of the data of a second questionnaire study related to the presence and placement of *te*. In that subsection, I only presented the results of the test sentences in 1-2-3-order, but note that in this questionnaire, all six logically possible word orders (1-2-3, 1-3-2, 2-1-3, 2-3-1, 3-1-2, 3-2-1) of the three cluster types were tested (see Appendix B for the full list of test items). An in-dept statistical and theoretical analysis of these data still needs to be carried out, but I already want to discuss here that in general, there is a difference in which word orders are allowed in the non-finite three-verb clusters compared to their finite counterpart. This is explicitly tested in the questionnaire for cluster type I. The non-finite three-verb cluster is repeated again in (13). Its finite counterpart is given in (14), in which V1 is a finite verb.

(13) Anne zegt haar oma gisteren lang [***te* zijn**$_1$ **blijven**$_2$ **helpen**$_3$].
Anne says her grandmother yesterday long to be.INF remain.INF help.INF
'Anne says she stayed to help her mom for a long time yesterday.'

(14) Anne zegt dat ze haar oma gisteren lang [**is**$_1$ **blijven**$_2$ **helpen**$_3$].
Anne says that she her grandmother yesterday long is.FIN remain.INF help.INF
'Anne says she stayed to help her mom for a long time yesterday.'

The numbers of speakers who accept (i.e. rated the sentence with a 4 or a 5) a given word order of the non-finite version of cluster type I (13) and the numbers of speakers who accept the finite-version of this cluster type (14) are given in Table 6.2. Note that these numbers are not cumulative: many speakers allow more than one word order.

As can be seen from the table, the most striking difference in numbers of speakers who accepted a given word order in the non-finite versus the finite counterpart is the 2-1-3 order. This order is accepted by 99 speakers in the non-finite version, and only by 7 speakers in the finite version. The 2-1-3 order is generally seen as an impossible word order in finite three-verb clusters, or is taken to be very marginal or only to occur in very specific types of cluster in specific dialects (see among others Abels (2016), Wurmbrand (2017), Salzmann (2019b)). Interestingly, this order is even more frequent in the non-finite cluster types II and III. The numbers of speakers who accepted the different word orders of these cluster types are given in Table 6.2. Unfortunately, their finite

Word order	Non-finite version	Finite version
1-2-3	367	472
1-3-2	44	72
2-1-3	99	7
2-3-1	39	69
2-3-1	6	4
2-3-1	12	21

Table 6.1: Frequencies of accepted word orders in non-finite and finite version of cluster type I

counterparts were not tested, but as already mentioned above, there is a large amount of empirical evidence that the 2-1-3 order is not accepted in finite three-verb clusters in Dutch (see also Barbiers et al. (2008, 2018), Dros-Hendriks (2018)).

Word order	Cluster type II	Cluster type III
1-2-3	458	493
1-3-2	103	47
2-1-3	217	176
2-3-1	23	57
2-3-1	85	13
2-3-1	18	26

Table 6.2: Frequencies of accepted word orders in cluster type II and III (non-finite)

It thus seems that the 2-1-3 order is accepted by a substantial amount of speakers in non-finite three-verb clusters. Future work is needed to understand why this is so, and to furthermore investigate the effect of the different word orders on the presence and placement of *te* in non-finite verb clusters more specifically.

Case study II

Semi-lexicality in Afrikaans pseudocoordination

CHAPTER 7

Introduction

> *Toe oom Hanges deur gellings water en wyn sy oë begin oopmaak, toe buk die dominee oor hom en hy vra: "Broer Hanges, sê my: nóú weet jy darem seker van beter?" Oom Hanges swyg en spu net see en wyn uit. "Sê my, broer," hou Ou-dominee vol: "Wat het daar by jou opgekom toe jy nou daar wegsink en jy sien jou laaste uur het geslaan?" Waarop oom Hanges, steeds spoegende, half-orent kom en sy woord spreek: "Nee, Dominee," seg hy, "toe ek nou dag dis vir goed klaar met my in die branders, toe kom ek tot 'n diep insig." "Ja, nè?" vra Dominee triomfantelik. "En dié insig is, broer?" "Dié insig," sê oom Hanges, "is, Dominee: dat dit darem 'n dêm mal ding is om te doen om te wil loop staan en Engeland toe swem as 'n man nie eens kan Engels praat nie."*
>
> \- Mal en ander stories, *Andé P. Brink*

In the first case study, one of the semi-lexically used verbs that was investigated in Dutch was the posture verb *zitten* 'sit', which can bring about a progressive reading of the lexical verb. An example is given here again in (1).

(1) Ik heb ***zitten*** **(te) lezen**.
I have sit to read
'I've been reading.'

I argued that this semi-lexically used verb is part of a small group of similarly behaving verbs, namely the motion verb *loop* 'walk', and the two other cardinal posture verbs, *staan* 'stand', and *liggen* 'lie'. All these verbs are semi-lexical, and have the same underlying structure in which the semi-lexically used verb is a root merged in the functional domain of the lexical verb (see subsection 5.4.4 for details).

In this case study, I look at the Afrikaans cognates of these verbs, namely *loop* 'walk', *sit* 'sit', *staan* 'stand' and *lê* 'lie'. More specifically, I investigate morphosyntactic variation and optionality in constructions in which these semi-lexically used verbs combine with a lexical verb. In contrast to Dutch, these four verbs, when combined with a lexical verb, appear in a so-called *pseudo-coordination* construction (henceforth PC construction).[1] That is, these verbs (henceforth PC verbs) are combined with the lexical verb by *en* 'and', as illustrated in (2)–(5).[2,3]

(2) Ek het gister baie **loop *en* praat**.
I have yesterday a.lot walk and talk
'I've (walked and) talked a lot yesterday.'

(3) Ek het die hele middag **sit *en* lees**.
I have the entire afternoon sit and read
'I've (sat and) read the entire afternoon.'

(4) Ek het vir ure met my ma op die telefoon **staan *en* praat**.
I have for hours with my mom at the telephone stand and talk
'I've (stood and) talked with my mom on the phone for hours.'

(5) Ek het die hele naweek **lê *en* slaap**.
I have the entire weekend lie and sleep
'I've (lied down and) slept the entire weekend.'

Even though the use of *en* can give the impression that the PC verb and the lexical verb are in a regular coordination configuration, these examples are cases of *pseudo*coordination. A clear indication of this is the fact that extraction of an

[1] Another label for this construction in Dutch is *verbale hendiadis* 'verbal hendiadys', from the original meaning of *hendiadys* as 'one by means of two' (Roberge 1994:45).

[2] In the translations of examples of PC constructions throughout this chapter, I give the translation of the PC verb in brackets. I do this following other work on Afrikaans PC constructions, such as Roberge (1994) and De Vos (2005). Their intention is to indicate that in many cases the lexical semantics of the PC verb is still strongly present in the meaning of the entire PC construction. In other words, in many cases the PC verb contributes both its own lexical semantics as well as a progressive or durative interpretation. Note that for some speakers this interpretation can also be of a different nature. For example, the PC verb *loop* 'walk' can also contribute an andative interpretation, and for some speakers also an inchoative one. Furthermore, *loop* can even be used to indicate an immediate future (Du Toit 1905:96) in Orange River Afrikaans. *Staan* 'stand' can also indicate anterior aspect (Biberauer 2019b), and all PC verbs can also contribute a modal rather than a progressive interpretation in imperative constructions, see Breed (2017) for examples and discussion. In this case study, I focus on the progressive/durative interpretation of PC verbs, and additionally on the possible andative interpretation of the PC verb *loop*.

[3] The examples given in (2)–(5) are created by a native speaker of Afrikaans (Theresa Biberauer) for the questionnaire of this case study, see subsection 9.2.1.

argument of one of the verbs is possible, which in the case of a true coordination would constitute a clear violation of the Coordinate Structure Constraint (Ross 1967). The fact that such extraction is possible in PC constructions is illustrated for the PC verb *sit* 'sit' in (6), but note that it holds for all PC verbs.

(6) Wat **sit** Jan waarskynlik **en** **eet** t_{wat}?
what sit Jan probably and eat
'What is Jan probably (sitting and) eating?' (De Vos 2005:135)

Another test for pseudocoordination is the possibility of substituting *en* by another conjunction (De Vos 2005). For example, substituting *en* 'and' by *of* 'or' leads to strong ungrammaticality in PC constructions. This is illustrated for *sit* 'sit' in (7). This is another indication that we are dealing with pseudocoordination rather than real coordination (see De Vos (2005) for more pseudocoordination tests applied on Afrikaans PC constructions).

(7) Wat **sit** Jan { ***en*** / ****of*** } **lees**?
what sit Jan and / or read
'What is Jan (sitting and/*or) reading?' (De Vos 2005:146)

Different aspects of Afrikaans PC have been discussed in detail by amongst others Kocks (1951), Ponelis (1993), Roberge (1994), Robbers (1997), De Vos (2001, 2005), Breed (2012), Breed (2017), Breed et al. (2017), Biberauer and Vikner (2017) and Biberauer (2019b). This case study contributes to the discussion in two ways. First, by presenting new data from both a corpus study and a questionnaire, in which the focus lies on morphosyntactic variation and optionality in PC constructions. Second, by proposing a formal analysis of these phenomena within the approach to semi-lexicality outlined in this dissertation. Specifically, this case study focusses on three ways in which PC constructions can display morphosyntactic optionality: (i) the presence/absence of *en*, (ii) the presence/absence and position of the participial marker *ge-*, and (iii) the behaviour of PC constructions under Verb Second. I illustrate each type of optionality in turn. Regarding the presence or absence of *en* in Afrikaans PC constructions, De Vos (2005) and Biberauer (2017c) have already pointed out that with the motion verb *loop* 'walk', *en* is often left out. An example of such a PC construction in which *en* is optional is given in (8).

(8) Daar **loop** hy alweer allerhande stories (***en***) **vertel**!
there walk he again all.kinds stories and tell
'There he's going around telling all kinds of stories again!' (Biberauer 2019b:6)

Regarding the optionality of *en* with the PC verb *loop*, it has also been noted that this is often related to the interpretation of this verb. Specifically, we see that it can be used to bring about a progressive/durative interpretation, such as the one in (2), or an andative one (De Vos 2001, 2005, Biberauer 2017c). I take andative elements to be elements 'that express the meaning of "go" or "come"'

(Cinque 2006:49). In the case of Afrikaans *loop*, the andative use corresponds to the meaning of "go", i.e. corresponding to movement away from the deictic centre (Ross 2016). An example of this use of *loop* is given in (9).

(9) Sal jy net gou vir my die groente **loop** (***en***) **bring**?
shall you just quickly for me the vegetables walk and bring
'Can you just quickly go and get me vegetables?' (Biberauer 2019b:11)

Specifically with the andative use of *loop*, speakers tend to leave out *en* (De Vos 2001, 2005, Biberauer 2019b), and less so with the progressive/durative use. This seems to be the case for most speakers of Afrikaans. In addition to this type of optionality, *en* has also been argued to be optional when following the PC verb *staan*, but only in specific varieties of Afrikaans, such as Orange River Afrikaans (De Vos 2005, Roberge 1994, Biberauer 2019b).[4] In this case study, the optionality of *en* in PC constructions is investigated with both a corpus study and a questionnaire.

The second aspect of variation and optionality in PC constructions concerns the presence or absence of the past participle marker *ge-* when the PC construction is embedded under a perfective auxiliary. In Afrikaans PC constructions, *ge-* seems to be truly optional (Roberge 1994, Robbers 1997, De Vos 2005). This is illustrated for the PC verb *sit* in (10).

(10) Hulle het op die stoep (***ge***-)**sit en rook**.
they have on the porch GE-sit and smoke
'They were (sitting and) smoking on the porch.' (Roberge 1994:46)

This type of morphosyntactic optionality is again investigated for all PC verbs, by means of a corpus study and a questionnaire. In addition to being optional, *ge-* can also appear on the lexical verb rather than the PC verb in certain varieties of Afrikaans (Roberge 1994, De Vos 2005), again most notably in Orange River Afrikaans. An example is given in (11).

(11) Dáár had [hy] **loop *ge*-kamp**.
there had he walk GE-fight
'There, he went to fight.' (Van Rensburg 1987:81)

As can be seen in the example, in this PC construction, *en* is absent. Whether the placement of *ge-* on the lexical verb is possible in other varieties than Orange River Afrikaans, and whether it correlates with the absence of *en* in these constructions, is investigated in the questionnaire of this case study.

The third aspect of variation and optionality in PC constructions that is investigated in this case study concerns their interaction with Verb Second (henceforth 'V2'). Like most other Germanic languages, Afrikaans exhibits V2. That is, in root clauses, the finite verb appears in the second position of the

[4] Orange River Afrikaans is a cover-term for a group of Afrikaans varieties that are spoken in the north-west Cape, in communities living along the Orange River (Theresa Biberauer p.c.).

clause, preceded by one other (phrasal or simplex) constituent. In contrast to the other Germanic languages exhibiting V2, however, in some contexts Afrikaans allows the finite and the lexical verb to be moved together into the V2 position, a phenomenon known as *Quirky Verb Second* (henceforth 'quirky V2') (De Vos 2005).[5] Both 'normal V2', in which only the finite verb is moved to the V2 position, and quirky V2, are illustrated for the PC verb *lê* 'lie' in (12).

(12) a. Hy **lê** die heeldag na die wolke **en kyk**.
He lies the whole.day at the clouds and look

b. Hy **lê en kyk** die heeldag na die wolke.
He lies and watch the whole.day at the clouds
'He is (lying and) looking at the clouds the entire day.' (Robbers 1997:65)

This type of optionality is investigated in the questionnaire as well.

In this case study, I argue that all Afrikaans PC verbs are semi-lexical. The andative use of *loop* is at the second stage of semi-lexicality, while all other PC verbs are in the first stage. In addition, I show that (i) the optionality of *en* is due to the stage of semi-lexicality the PC verb is in, (ii) the optionality of *ge-* is due to the morphosyntactic status of this element and its role in the expression of past tense, and (iii) the alternation between normal and quirky V2 is due to two different underlying structures.

This part is structured as follows. In chapter 8, I present the methodology and results of the corpus study. In chapter 9, I do the same for the questionnaire study. In chapter 10, based on these two studies, I present my analysis of morphosyntactic variation and optionality in Afrikaans PC constructions. In chapter 11, I discuss the findings of this case study, and conclude with a direction for future research.

[5]Note that in Afrikaans, due to the lack of verbal inflection, finite verbs cannot be distinguished morphologically from infinitival ones (a few exceptions notwithstanding, such as *is/wees* 'be' and possessive *het/hê* 'have').

CHAPTER 8

The corpus study

8.1 Introduction

This chapter deals with the methodology and results of the corpus study. In section 8.2, I present the methodology, which includes the description of the corpus and the execution of the data extraction. In section 8.3, I present the results of the corpus study.

8.2 Methodology

8.2.1 Description of the corpus

The data presented in this chapter are taken from the *Korpusportaal* corpus (VivA 2016), which, at the time of the search, has a total of 85 million words.[1] The corpus contains printed and electronic texts of standard and regional Afrikaans, including texts written to be spoken (radio broadcasts) of various registers and genres. The corpus is a collection of seven subcorpora, namely the *NCHLT-Afrikaanse korpus 1.0* (3.2 million words), the *NWU/Maroela Media-korpus 1.0* (800,000 words), the *NWU/Lapa-korpus 1.0* (12 million words), the *PUK/Protea Boekhuis-korpus 2.0* (12 million words), the *RSG-nuuskorpus 2.0* (20.5 million words), the *Taalkommissie-korpus 1.1* (47 million words) and the *Watkykjy.co.za-korpus 1.0* (around 1 million words).

[1] This was the corpus size at the time at which I performed my corpus search, namely March 2017.

8.2.2 Data extraction

In the corpus study, I only extracted PC constructions that were embedded under the perfective auxiliary *het* 'have'. In Afrikaans, in embedded clauses, the only possible word order in PC constructions embedded under such an auxiliary is V2-*en*-V3-V1) (Wurmbrand 2017), in which V2 is the PC verb, V3 the lexical verb, and V1 the perfective auxiliary. An example is given in (1).

(1) ... dat ek die hele middag **sit**$_2$ ***en*** **lees**$_3$ **het**$_1$.
... that I the entire afternoon sit and read have
'... that I've been (sitting and) reading the entire afternoon.'

In root clauses, the perfective auxiliary *het* (V1) is in V2 position, and the PC construction is in clause final position. An example is given in (2).[2]

(2) Ek **het**$_1$ die hele middag **sit**$_2$ ***en*** **lees**$_3$.
I have the entire afternoon sit and read
'I've been (sitting and) reading the entire afternoon.'

Both configurations were extracted from the corpus, to make sure all uses of PC constructions embedded under a perfective auxiliary were found. For both configurations, the optionality of the past participle marker *ge-* on the PC verb as well as the optionality of *en* was investigated. In the queries, the PC verb was always searched for as 'lemma', so that the query was able to return PC constructions in which the PC verb appears either with or without *ge-*. The same holds for *en*: this morpheme was indicated as optional in the queries (by placing a '?' after [word = "en"]). The queries that were used for each PC verb are given below. The (a) example always gives the query for embedded clauses, and the (b) example the query for root clauses. For this latter type, the $[]\{0,5\}$ part of the query ensures that it returns all sentences in which there are minimally zero and maximally 5 tokens between the finite verb in V2 position, and the start of the PC construction.

(3) a. [*lemma* = "*loop*"][*word* = "*en*"]?[*pos_head* = "*ww*"] [*word* = "*het*"]
b. [*word* = "*het*"] $[]\{0,5\}$ [*lemma* = "*loop*"][*word* = "*en*"]?[*pos_head* = "*ww*"]

(4) a. [*lemma* = "*sit*"][*word* = "*en*"]?[*pos_head* = "*ww*"] [*word* = "*het*"]
b. [*word* = "*het*"] $[]\{0,5\}$ [*lemma* = "*sit*"][*word* = "*en*"]?[*pos_head* = "*ww*"]

[2] Recall that in the introduction of this case study (chapter 7), I have introduced the notion of 'Quirky V2', in which the entire PC construction appears in V2 position. Note that Quirky V2 is excluded when the finite verb is an auxiliary or modal verb (De Vos 2005), which I discuss in more detail in section 10.2. For now, it suffices to observe that a Quirky V2 variant of (2) – in which more verbs than auxiliary *het* are placed in V2 position – is ungrammatical.

(5) a. $[lemma = \text{“}staan\text{”}][word = \text{“}en\text{”}]?[pos\_head = \text{“}ww\text{”}]\ [word = \text{“}het\text{”}]$
b. $[word = \text{“}het\text{”}]\ []\{0,5\}\ [lemma = \text{“}staan\text{”}][word = \text{“}en\text{”}]?[pos\_head = \text{“}ww\text{”}]$

(6) a. $[lemma = \text{“}le\text{”}][word = \text{“}en\text{”}]?[pos\_head = \text{“}ww\text{”}]\ [word = \text{“}het\text{”}]$
b. $[word = \text{“}het\text{”}]\ []\{0,5\}\ [lemma = \text{“}le\text{”}][word = \text{“}en\text{”}]?[pos\_head = \text{“}ww\text{”}]$

As mentioned in the introduction of this case study (chapter 7), this case study investigates both the progressive/durative and the andative use of the PC verb *loop*. In the questionnaire study, separate test items are used to distinguish between these two different interpretations. In a corpus study, however, the aspectual interpretation of *loop* in the sentences cannot be manipulated, and needs to be annotated by native speakers. For all hits that involved PC constructions with *loop*, two native speakers of Afrikaans (one 28-year old male, and one 46-year old female) were asked to annotate them with respect to the interpretation brought about by the PC constructions. The instruction text for the annotators can be found in Appendix C.[3]

The optionality of normal versus quirky V2 was not investigated in this corpus study. Therefore, this type of optionality is not discussed in the next section. I return to this phenomenon when discussing the questionnaire study, i.e. in chapter 9.

8.3 Results of the corpus study

In this section, I present the results of the corpus study. Table 8.1 contains the total number of hits per PC verb that was returned by the queries.[4]

Verb	Total
Loop ‘walk’	109 (100%)
Sit ‘sit’	455 (100%)
Staan ‘staan’	346 (100%)
Lê ‘lie’	249 (100%)

Table 8.1: Total number of hits per PC verb

The highest number of hits are PC constructions with *sit*, followed by *staan*, then *lê*, and lastly *loop*. Let us now look at the morphological form of the PC

[3]Note that I am aware that ideally the set of PC constructions with *loop* would have been annotated by a larger number of native speakers of Afrikaans. However, as the two uses of *loop* are investigated in the questionnaire study as well, which was filled out by 201 native speakers of Afrikaans – see section 1.5, I take the results of the annotations to serve as additional data to the data of the questionnaire study, and are thus not used in isolation.

[4]The total number of hits are the hits of embedded and root clauses combined.

verbs in these examples, i.e. does it bear the past participle marker *ge-* or not? The results are given in Table 8.2.

Verb	Without *ge-*	With *ge-*	Total
Loop 'walk'	83 (76,1%)	28 (23,9%)	109 (100%)
Sit 'sit'	220 (48,4%)	235 (51,6%)	455 (100%)
Staan 'staan'	155 (44,8%)	191 (55,2%)	346 (100%)
Lê 'lie'	113 (45,4%)	136 (54,6%)	249 (100%)

Table 8.2: Morphological form of the PC verb

As can be seen in the table, there is quite a difference in frequency between the presence of *ge-* for the PC verb *loop* on the one hand, and for the posture verbs on the other. Whereas the former appears much more frequently without *ge-* (74,8%), all posture verbs appear with comparable frequencies with and without *ge-*. T-tests show that the results for *loop* differ significantly from all posture verbs (p-value <0.01 in all cases), whereas there is no significant difference among the three posture verbs. The t- and p-value for all T-tests are given in Table 8.3.

	Loop	*Sit*	*Staan*	*Lê*
Loop	–	–	–	–
Sit	t=6.3, p-value<0.01*	–	–	–
Staan	t=6.4, p-value<0.01*	t=0.4, p-value=0.68	–	–
Lê	t=5.9, p-value<0.01*	t=0.2, p-value=0.82	t=-0.1, p=0.89	–

Table 8.3: T-test results concerning the presence/absence of *ge-* on the PC verbs

Let us now look at the presence or absence of *en* in the PC constructions extracted from the corpus. The frequencies per PC verb are given in Table 8.4. The data show that *en* can only remain absent with the PC verb *loop*. The posture PC verbs all show zero cases of *en*-absence. Furthermore, with the PC verb *loop*, *en* is much more frequently absent than present: in 78,4% versus 21,6% of the cases. Given that with *loop* both *ge-* and *en* can either be present or absent, let us also look at these two factors combined. The distribution of the two factors for *loop* is given in Table 8.5. The distribution is as follows. The most frequent combination of the two factors is when both *ge-* and *en* are absent, i.e. '*loop* V'. The second most frequent combination, is when they are both present, i.e. '*geloop en* V'. The other two combinations, with *ge-* but without *en* (i.e. '*geloop* V'), and with *en* but without *ge-* (i.e. '*loop* en V'), are very infrequent.

Recall from section 8.3 that the set of *loop* PC constructions was annotated

Verb	*En* present	*En* absent	Total
Loop 'walk'	24 (21,6%)	85 (78,4%)	109 (100%)
Sit 'sit'	455 (100%)	0 (0%)	455 (100%)
Staan 'staan'	346 (100%)	0 (0%)	346 (100%)
Lê 'lie'	249 (100%)	0 (0%)	249 (100%)

Table 8.4: Presence/absence of *en* 'and' in the PC constructions

	Without *ge*-	With *ge*-	Total
En present	7 (6,4%)	17 (15,6%)	24 (22%)
En absent	76 (69,7%)	9 (8,3%)	85 (78%)
Grant total			109 (100%)

Table 8.5: Presence/absence of *ge*- and *en* with PC verb *loop*

by two native speakers of Afrikaans with respect to the type of aspect expressed by *loop*: andative or progressive. The frequencies are given in Table 8.6. The 'Unclear' category contains the cases in which the annotators were unsure about the interpretation of the sentence, or when they had conflicting judgments. The data show that the most common category is 'Andative' in the annotated data set, but it is worth pointing out that the category 'Unclear' is also quite large.

Type of aspect	total
Andative	44 (40,4%)
Progressive	25 (22,9%)
Unclear	40 (36,7%)
Total	109 (100%)

Table 8.6: Type of aspect in *loop* PCs

Let us now look at the interaction between the three factors discussed so far, i.e. the presence or absence of *ge*-, the presence or absence of *en*, and the two types of aspect. The frequencies are given in Table 8.7.

The data show that when *loop* expresses andative aspect, the most frequent configuration is one in which it occurs without *ge*-, and in which *en* is also absent (i.e. '*loop* V'). When *loop* expresses progressive aspect, two configurations are equally frequent, namely the one in which both *ge*- and *en* are present (i.e. '*geloop en* V'), and the one in which both are absent (i.e. '*loop* V'). In other words, in the case of the andative use of *loop*, one configuration is strongly preferred: '*loop* V'. The progressive use of *loop* shows more morphosyntactic

	en present		*en* absent		
Aspect	*ge*- absent	*ge*- present	*ge*- absent	*ge*- present	Total
Andative	2 (4,5%)	1 (2,3%)	38 (86,4%)	3 (6,8%)	44 (100%)
Progressive	2 (8,0%)	11 (44,0%)	11 (44,0%)	1 (4,0%)	25 (100%)
Unclear	3 (7,5%)	6 (15,0%)	27 (67,5%)	4 (10%)	40 (100%)
Total					109 (100%)

Table 8.7: Aspect combined with the presence/absence of *ge*- and *en* in *loop* PC constructions

optionality in PC constructions.

Summing up, this corpus study has revealed the following patterns regarding the morphosyntax of Afrikaans PC constructions. First, we have seen that the PC verb *loop* behaves significantly differently from the posture PC verbs. With respect to the presence or absence of *ge*-, we have seen that *loop* shows very high frequencies of *ge*-drop in PC constructions, whereas for the posture verbs, the frequencies for the presence vs. absence of *ge*- are more or less equal. Regarding the presence or absence of *en*, we have seen that only *loop* PC constructions can occur without *en*, and furthermore, that the absence of *en* is clearly the more frequent option. When examining the interaction between these two factors, we see that for the PC verb *loop* the most frequent configuration is the one in which both *ge*- and *en* are absent. With respect to the type of aspect expressed by the PC verb *loop*, the data set seems to contain more andative uses than progressive ones, but there is also a high number of unclear cases. As for the interaction between the type of aspect and the presence or absence of *ge* and *en*, we have seen that when *loop* expresses andative aspect, the most frequent configuration by far is the one in which both *ge*- and *en* are absent (i.e. '*loop* V'), whereas when *loop* expresses progressive aspect, two configurations are equally frequent, namely the one in which both *ge*- and *en* are absent (i.e. '*loop* V'), and the one in which they are both present (i.e. '*geloop en* V').

CHAPTER 9

The questionnaire study

9.1 Introduction

In this chapter, I present the methodology and the results of a questionnaire study into Afrikaans PC constructions. Section 9.2 discusses the methodology, in particular the design of the questionnaire, the task and procedure, and the participants. In section 9.3, the results are presented and discussed.

9.2 Methodology

9.2.1 Design

The design, task and procedure of this questionnaire were modelled after those of the Dutch questionnaire presented in case study I. Most decisions concerning the design and procedure were based on Schütze (2016).

The goal of this questionnaire was twofold. First, it investigates variation in the verb form of the PC verbs when embedded under a perfective auxiliary, i.e. with or without *ge-*, and with or without *en*. Second, it investigates the optionality of normal V2 versus quirky V2 in sentences in which the PC verb is the finite verb of the clause. In these cases, the presence or absence of *en* is also investigated. Given that *loop* 'walk' can convey both progressive and andative aspect, both types of PC constructions were tested. In addition, all three posture verbs were tested as well. The test sentences in which the PC verbs are embedded under the perfective auxiliary *het* 'has' are given in (1)–(5).

(1) *Loop*$_{andative}$ PC – embedded condition[1]

a. Paul sê dat Lisa verlede week 'n splinternuwe motor **loop en koop** het.
Paul says that Lisa last week a brand.new car walk and buy has

b. Paul sê dat Lisa verlede week 'n splinternuwe motor **geloop en koop** het.
Paul says that Lisa last week a brand.new car GE-walk and buy has

c. Paul sê dat Lisa verlede week 'n splinternuwe motor **loop koop** het.
Paul says that Lisa last week a brand.new car walk buy has

d. Paul sê dat Lisa verlede week 'n splinternuwe motor **geloop koop** het.
Paul says that Lisa last week a brand.new car GE-walk buy has
'Paul says Lisa went and bought a brand new car last week.'

(2) *Loop*$_{progressive}$ PC – embedded condition

a. Steve sê dat Cornelia gisteraand baie **loop en praat** het.
Steve says that Cornelia yesterday.night a.lot walk and talk has

b. Steve sê dat Cornelia gisteraand baie **geloop en praat** het.
Steve says that Cornelia yesterday.night a.lot GE-walk and talk has

c. Steve sê dat Cornelia gisteraand baie **loop praat** het.
Steve says that Cornelia yesterday.night a.lot walk talk has

d. Steve sê dat Cornelia gisteraand baie **geloop praat** het.
Steve says that Cornelia yesterday.night a.lot GE-walk talk has
'Steve says that Cornelia has been (walking and) talking a lot yesterday night.'

(3) *Sit* PC – embedded condition

a. Simon sê dat Thomas die hele middag **sit en lees** het.
Simon says that Thomas the entire afternoon sit and read has

b. Simon sê dat Thomas die hele middag **gesit en lees** het.
Simon says that Thomas the entire afternoon GE-sit and read has

[1] We can see that *loop* is used as an andative by the English translation 'went and' (Cinque 2006:49).

c. Simon sê dat Thomas die hele middag **sit lees** het.
Simon says that Thomas the entire afternoon sit read has

d. Simon sê dat Thomas die hele middag **gesit lees** het.
Simon says that Thomas the entire afternoon GE-sit read has
'Simon says Thomas has been (sitting and) reading the entire afternoon.'

(4) *Staan* PC – embedded condition

a. Susan sê dat Elsa vir ure met haar ma op die telefoon **staan en praat** het.
Susan says that Elsa for hours with her mom at the telephone stand and talk has

b. Susan sê dat Elsa vir ure met haar ma op die telefoon **gestaan en praat** het.
Susan says that Elsa for hours with her mom at the telephone GE-stand and talk has

c. Susan sê dat Elsa vir ure met haar ma op die telefoon **staan praat** het.
Susan says that Elsa for hours with her mom at the telephone stand talk has

d. Susan sê dat Elsa vir ure met haar ma op die telefoon **gestaan praat** het.
Susan says that Elsa for hours with her mom at the telephone GE-stand talk has
'Susan says that Elsa has been (standing and) talking with her mom on the phone for hours.'

(5) *Lê* PC – embedded condition

a. Eric sê dat Michael die hele naweek **lê en slaap** het.
Eric says that Michael the entire weekend lie and sleep has

b. Eric sê dat Michael die hele naweek **gelê en slaap** het.
Eric says that Michael the entire weekend GE-lie and sleep has

c. Eric sê dat Michael die hele naweek **lê slaap** het.
Eric says that Michael the entire weekend lie sleep has

d. Eric sê dat Michael die hele naweek **gelê slaap** het.
Eric says that Michael the entire weekend GE-lie sleep has
'Eric says that Michael has been (lying and) sleeping the entire weekend.'

Note that for all test items, there were four versions, of the following abstract format:

1. PC verb *en* lexical-verb
2. *Ge*-PC verb *en* lexical-verb
3. PC verb lexical-verb

4. *Ge*-PC verb lexical-verb

In addition to these four versions, in the case of *loop*$_{andative}$ and *sit*, a fifth and sixth version were added, namely one in which *ge*- appears on the lexical verb rather than the PC verb, both in a PC construction with and one without *en*. These two additional versions are given in (6) for *loop*$_{andative}$ and in (7) for *sit*.[2]

(6) *Loop*$_{andative}$ PC - *ge*- on lexical verb

a. Paul sê dat Lisa verlede week 'n splinternuwe motor **loop en** **ge-koop** het.
   Paul says that Lisa last week a splinter.new car walk and GE-buy has

b. Paul sê dat Lisa verlede week 'n splinternuwe motor **loop** **ge-koop** het.
   Paul says that Lisa last week a splinter.new car walk GE-buy has
   'Paul says Lisa went and bought a brand new car last week.'

(7) *Sit* PC - *ge*- on lexical verb

a. Simon sê dat Thomas die hele middag **sit en ge-lees** het.
   Simon says that Thomas the entire afternoon sit and GE-read has

b. Simon sê dat Thomas die hele middag **sit ge-lees** het.
   Simon says that Thomas the entire afternoon sit GE-read has
   'Simon says Thomas has been (sitting and) reading the entire afternoon.'

The test sentences per PC verb that were used to investigate the optionality of normal and quirky V2 are given in (8)-(11). Note that only progressive *loop* was used in this part of the questionnaire, and not andative. This decision was made in order to limit the total length of the questionnaire.

(8) *Loop*$_{progressive}$ PC – V2 condition

a. Hoekom **loop** Jan heeldag piesangs **en eet**?
   why walk.FIN Jan all.day bananas and eat

b. Hoekom **loop en eet** Jan heeldag piesangs?
   why walk.FIN and eat Jan all.day bananas

c. Hoekom **loop** Jan heeldag piesangs **eet**?
   why walk.FIN Jan all.day bananas eat

d. Hoekom **loop eet** Jan heeldag piesangs?
   why walk.FIN eat Jan all.day bananas

[2] In order to prevent the questionnaire from becoming too long, these two versions were only tested for one of the two *loop* PC verbs and for one of the posture PC verbs, rather than for all PC verbs.

'Why is Jan (walking and) eating bananas all day long?'

(9) *Sit* PC – V2 condition

a. Hoekom **sit** Lisa heeldag die koerant **en lees**?
why sit.FIN Lisa all.day the newspaper and read

b. Hoekom **sit en lees** Lisa heeldag die koerant?
why sit.FIN and read Lisa all.day the newspaper

c. Hoekom **sit** Lisa heeldag die koerant **lees**?
why sit.FIN Lisa all.day the newspaper read

d. Hoekom **sit lees** Lisa heeldag die koerant?
why sit.FIN read Lisa all.day the newspaper
'Why is Lisa (sitting and) reading the newspaper all day long?'

(10) *Staan* PC – V2 condition

a. Hoekom **staan** Thomas heeldag sy ken **en vryf**?
why stand.FIN Thomas all.day his chin and rub

b. Hoekom **staan en vryf** Thomas heeldag sy ken?
why stand.FIN and rub Thomas all.day his chin

c. Hoekom **staan** Thomas heeldag sy ken **vryf**?
why stand.FIN Thomas all.day his chin rub

d. Hoekom **staan vryf** Thomas heeldag sy ken?
why stand.FIN rub Thomas all.day his chin
'Why is Thomas (standing and) rubbing his chin all day long?'

(11) *Lê* PC – V2 condition

a. Hoekom **lê** Mark heeldag na die wolke **en kyk**?
why lie.FIN Mark all.day at the clouds and look

b. Hoekom **lê en kyk** Mark heeldag na die wolke?
why lie.FIN and look Mark all.day at the clouds

c. Hoekom **lê** Mark heeldag na die wolke **kyk**?
why lie.FIN Mark all.day at the clouds look

d. Hoekom **lê kyk** Mark heeldag na die wolke?
why lie.FIN look Mark all.day at the clouds
'Why is Mark (lying and) looking at the clouds all day?'

Once again, four versions were tested for each test item. They are abstractly represented below:

1. PC verb …… *en* lexical-verb
2. PC verb *en* lexical-verb ……
3. PC verb …… lexical-verb
4. PC verb lexical-verb ……

Participants were asked an additional question, if they rated one of the test items in (8)–(11) with a 4 or a 5.[3] This question is given in (12)–(15) for each of the PC verbs.

(12) Is Jan besig om te loop?
is Jan busy for to walk
'Is Jan walking?'

(13) Is Lisa besig om te sit?
is Lisa busy for to sit
'Is Lisa sitting?

(14) Is Thomas besig om te staan?
is Thomas busy for to stand
'Is Thomas standing?'

(15) Is Mark besig om te lê?
is Mark busy for to lie
'Is Mark lying?'

The participants were asked to answer these questions with 'yes', 'no', or 'uncertain'.

Besides these two groups of test items, other relevant sentence types were also included in the questionnaire. One group of test items checks whether the PC verbs can combine with a lexical verb the semantics of which is incompatible with the lexical semantics of the PC verb. The relevant test items are given in (16)–(19). For each test item, two versions were tested, one without *ge-* on the PC verb, and one with *ge-* on the PC verb. Only the progressive use of *loop* was tested in order to limit the length of the questionnaire.

(16) $Loop_{progressive}$ PC – incompatible lexical verb

a. Eva sê dat Pieter die laaste tyd die hele middag **loop**
Eva says that Pieter the last period the entire afternoon walk
en swem het.
and swim het

b. Eva sê dat Pieter die laaste tyd die hele middag
Eva says that Pieter the last period the entire afternoon
geloop en swem het.
GE-walk and swim het
'Eva says that lately, Pieter has been swimming entire afternoon.'

(17) *Sit* PC – incompatible lexical verb

a. Sara sê dat Thomas die laaste tyd die hele middag **sit**
Sara says that Thomas the last period the entire afternoon sit
en swem het.
and swim het

[3]See subsection 4.3.2 of case study I on the ratings 4 and 5 being interpreted as the sentence being grammatical for a given speaker.

b. Sara sê dat Thomas die laaste tyd die hele middag
Sara says that Thomas the last period the entire afternoon
gesit en swem het.
GE-sit and swim het
'Sara says that lately, Thomas has been swimming entire afternoon.'

(18) *Staan* PC – incompatible lexical verb

a. Laura sê dat Pieter die laaste tyd die hele middag
Laura says that Pieter the last period the entire afternoon
staan en swem het.
stand and swim het

b. Laura sê dat Pieter die laaste tyd die hele middag
Laura says that Pieter the last period the entire afternoon
gestaan en swem het.
GE-stand and swim het
'Laura says that lately, Pieter has been swimming entire afternoon.'

(19) *Lê* PC – incompatible lexical verb

a. Ann sê dat Thomas die laaste tyd die hele middag **lê**
Ann says that Thomas the last period the entire afternoon lie
en swem het.
and swim het

b. Ann sê dat Thomas die laaste tyd die hele middag
Ann says that Thomas the last period the entire afternoon
gelê en swem het.
GE-lie and swim het
'Ann says that lately, Thomas has been swimming entire afternoon.'

The lexical verb of all these sentences is *swim* 'swim', which is generally incompatible with walking, or with a seated, standing or lying position.

The entire questionnaire consisted of 63 test items and 12 filler items, for a total of 75 sentences.[4] The items of the questionnaire were presented in randomised order. Before the actual questionnaire started, the participants took part in a practice round, which consisted of four practice items of different degrees of grammaticality. The entire questionnaire, including practice items and background questions, is given in Appendix D. The questionnaire was created in Limesurvey©.

[4]Note that for reasons of space not all 63 test items will be discussed in this chapter. See Appendix D for the full questionnaire.

9.2.2 Task and procedure

A Judgment task was used. Participants were asked to rate sentences using a five point Likert scale. For each sentence, they were asked to say it out loud, in order to prompt them to rate the sentence as if it was spoken, rather than rate a written sentence. They were asked to answer the following question: *'is this a possible sentence in Afrikaans as it is spoken in your immediate environment?'*. Below the question, the five point Likert scale was given. 5 marked as 'certainly' and 1 was marked as 'certainly not'. The values in between were given as 4, 3, and 2. In the instruction text accompanying the questionnaire, 4, 3 and 2 were defined as 'quite good, but not completely perfect', 'neither clearly good nor clearly bad' and 'quite bad, but not completely impossible' respectively. Participants could also assign 'I don't know' if they did not have any judgment about a sentence. For each sentence, a comment box was provided, in case participants wanted to comment on (their rating of) the sentence.

The participants were first presented with an instruction text. They were familiarised with the fact that native speakers often have a quite clear idea about what a possible sentence in their language is, and that they can even say so regarding the *structure* of a sentence. It was explained that they would be presented with a number of sentences, each of which they were (i) asked to say aloud, and (2) to judge with respect to the question introduced above. In the instruction text, 'immediate environment', the phrase that was used in the question participants had to answer, was defined as 'friends, family, city or town'. The full instruction text can be found in Appendix D. Prior to the actual questionnaire, all participants were given the same set of 4 practice sentences. This was done so as to give the participants the time to get used to the task, prior to them rating the first sentences of the actual questionnaire. After the participants had completed the questionnaire, they were asked to answer a number of questions to collect some background information. The full list of background questions can be found in Appendix D.

9.2.3 Participants

The participants were recruited via social media, personal networks, and via a blog post on the website *voertaal.nu.* 201 native speakers of Afrikaans completed the questionnaire.[5] Unlike in the Dutch questionnaire in case study I, participants were not excluded from analysis based on them living abroad at the time of the questionnaire, or having lived abroad for more than 10% of their lives. This decision was made because it would bring the total number of participants down by too much. The mean age of the participants was 49 years (*SD*=15.3, range=20-89). The gender division was 155 female, 46 male.

In Figure 9.1, the locations of the participants are presented on a map of South Africa. Note that it is not the case that one point represents only

[5]Four additional speakers completed the questionnaire, but were not included in the analysis, because their native language was Dutch rather than Afrikaans.

one speaker: there are many locations for which there are many representative speakers.[6] As can be seen on the map, there are no speakers from the Northern Cape province: the empty area in the north-west. In general, speakers from this region are very hard to reach with online questionnaires; data collection for this region usually involves going there and interviewing the speakers (Theresa Biberauer, p.c.). Unfortunately, my questionnaire did not reach speakers from this area area. Future fieldwork is thus needed in order to include these speakers in the study of morphosyntactic variation and optionality in PC constructions.

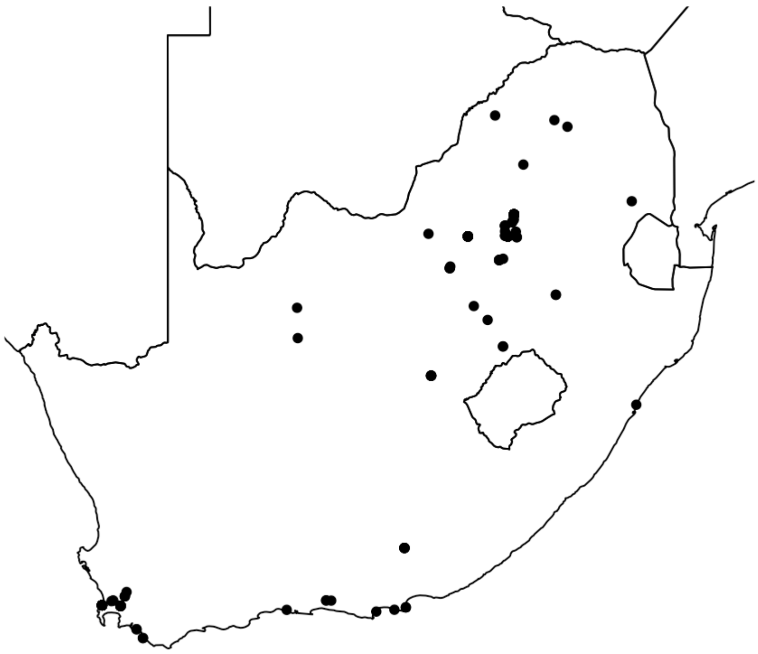


Figure 9.1: Locations of participants (case study II)

9.3 Results questionnaire study

9.3.1 Introduction

In this section, the results of the questionnaire study are presented and discussed. In subsections 9.3.2 and 9.3.3, I discuss the preliminaries and outline

[6] Note that the location of the participants who currently live abroad (N=27) is not indicated on this map.

the general findings. Subsections 9.3.4 and 9.3.5 deal with the optionality patterns and variation across the different PC verbs. Subsection 9.3.6 focuses on the semantic bleaching of the PC verbs. Subsection 9.3.7 discusses the results of exploratory statistical techniques (MCA, HC and CA) carried out on the raw data. Subsection 9.3.8 compares the results of the questionnaire study with those of the corpus study. Subsection 9.3.9 gives an overview of the findings that should be accounted for in the formal analysis.

9.3.2 Preliminaries

Let us start with the preliminaries, namely the number of speakers per type of PC construction that accepted that construction in general. Specifically, some speakers rejected all versions of the progressive use of *loop* in embedded PC constructions. These speakers were thus removed from the total number of speakers, so that we have a clear overview of the total number of speakers who accepted at least one version of a given PC construction. Recall from section 9.2.1 that two main types of PC constructions were tested: (i) PC constructions embedded under the perfective auxiliary *het*, and (ii) non-embedded PC constructions in which (a part of) the PC is in V2 position. The two types of constructions are illustrated again below for the PC verb *sit*. Recall that for the second type of PC, only the progressive use of *loop* was tested, in order to keep the length of the questionnaire within acceptable limits.

(20) a. Simon sê dat Thomas die hele middag **(ge)-sit** (**en**) **lees** ***het***.
Simon says that Thomas the entire afternoon GE-sit and read has
'Simon says Thomas has been (sitting and) reading the entire afternoon.'

b. Hoekom **sit** <**(en) lees**> Lisa heeldag die koerant <**(en) lees**>?
Why sit.FIN and read Lisa all.day the newspaper and read
'Why is Lisa (sitting and) reading the newspaper all day long?'

In Table 9.1, the numbers of speakers who accepted at least one version of the embedded PC constructions are given. Recall that the questionnaire was completed by a total of 201 native speakers of Afrikaans. 40 speakers did not accept any version of the $loop_{progressive}$ PCs, 35 speakers in the case of $loop_{andative}$ PCs, 3 for *sit* PCs, 9 for *staan* PCs, and 1 for *lê* PCs.

In Table 9.2, the numbers of speakers who accepted at least one version of the V2 PC constructions are given. 9 speakers did not accept any version with $loop_{progressive}$, 3 with *sit*, 5 with *staan* and 2 with *lê*.

As for the data in the previous case study (see 4.3.2), in this case study I have encoded the data as follows. For the largest part of the data presentation,

Type of PC	speakers who accepted at least one version
(Ge)loop$_{progressive}$	161
(Ge)loop$_{andative}$	166
(Ge)sit	198
(Ge)staan	192
(Ge)lê	200

Table 9.1: Number of speakers per PC type – embedded sentences

Type of PC	speakers who accepted at least one version
Loop$_{prog}$	192
Sit	198
Staan	196
Lê	199

Table 9.2: Number of speakers per PC type – V2 sentences

I have encoded the ratings 1, 2, and 3 as 0 (= ungrammatical) and the ratings 4 and 5 as 1 (= grammatical) for the largest part of the data presentation, i.e. in in 9.3.4–9.3.6 and in 9.3.7.4. The scale of 1-5 rating has been used as the input for the exploratory statistic method Multiple Correspondence Analysis (MCA), and for the Hierarchical Clustering (HC), for which I have used the MCA as input in this case study. These data are presented in 9.3.7.1–9.3.7.3.

9.3.3 General findings

Before going over the detailed results of the questionnaire study, I first want to highlight the main findings. The general findings of the questionnaire study are that all three variables under investigation show high degrees of optionality: (i) the presence or absence of the participial marker *ge*-, (ii) the presence or absence of *en*, and (iii) the type of V2-configuration (normal or quirky V2). Regarding the presence or absence of *ge*-, the data show that this element is highly optional for all speakers, in all types of embedded PC constructions. For example, most speakers allow both versions given in (21) of the embedded *loop*$_{progressive}$ PC construction.

(21) a. Steve sê dat Cornelia gisteraand baie **loop en praat** het.
Steve says that Cornelia yesterday.night a.lot walk and talk has

b. Steve sê dat Cornelia gisteraand baie **geloop en praat** het.
Steve says that Cornelia yesterday.night a.lot GE-walk and talk has
'Steve says that Cornelia has been (walking and) talking a lot yesterday night.'

Regarding the presence or absence of *en*, this element can be absent in all PC constructions, but with clearly different frequencies: *en* is most frequently absent in $loop_{andative}$ PC constructions, and much less so in the other PC constructions. Regarding the type of V2 construction, there too the choice between the two versions (normal versus quirky V2) was highly optional. Let us now turn to the exact frequencies of all these optionality patterns, discussed in detail in the next subsection.

9.3.4 Optionality patterns

Let us start with the optionality patterns in the embedded PC constructions. In Table 9.3, the frequencies for the presence or absence of *ge-* are given. Per PC verb, I present (i) the number of speakers who require *ge-* to appear on the PC verb ('Oblig. *ge-*), (ii) the number of speakers that allow *ge-* to be either present or absent on the PC verb ('Optional *ge-*'), and (iii) the number of speakers who require *ge-* to be absent ('Oblig. no *ge-*).

Verb	Oblig. *ge-*	Optional *ge-*	Oblig. no *ge-*	Total
$Loop_{progressive}$	39 (24,2%)	104 (64,6%)	18 (11,2%)	161 (100%)
$Loop_{andative}$	10 (6,0%)	123 (74,1%)	33 (19,9%)	166 (100%)
Sit	26 (13,1%)	168 (84,4%)	4 (2,5%)	198 (100%)
Staan	24 (12,5%)	159 (82,8%)	9 (4,7%)	192 (100%)
Lê	16 (8,0%)	182 (91,0%)	2 (1,0%)	200 (100%)

Table 9.3: Optionality of *ge-* per PC verb

As can be seen from the table, for all types of PC verbs the most frequent pattern is optional *ge-*, with the lowest percentage being 64,4% for the PC verb $loop_{progressive}$ and the highest percentage of 91,0% for the PC verb *lê*. A second pattern that can be observed is that $loop_{andative}$ is the only PC verb for which the percentage of obligatory *ge-* absence is higher than the one for obligatory *ge-* presence, respectively 19,9% versus 6,0%. For all other PC verbs we see the reverse pattern, namely that the percentage of obligatory *ge-* presence is higher than the percentage for its obligatory absence.

Recall from subsection 9.2.1 that for two of the five PC verb constructions, I also tested whether *ge-* could appear on the lexical verb rather than the PC verb. This was tested for one of the two uses of *loop*, namely $loop_{andative}$, and one of the posture verbs, namely *sit*. An example of this is repeated in (22).

(22) Simon sê dat Thomas die hele middag **sit en *ge*-lees** het.
Simon says that Thomas the entire afternoon sit and GE-read has
'Simon says Thomas has been (sitting and) reading the entire afternoon.'

The position of *ge*- on the lexical verb was tested both in versions with and without *en*. In Table 9.4, the frequencies of *ge*- on V3 (i.e. the lexical verb) are given. The column 'No *ge*- on V3' contains the numbers of speakers that do not allow *ge*- to appear on the lexical verb. In the column 'Opt. *ge*- on V3' the numbers of speakers are given who allow both the version with *ge*- on the lexical verb and the version without. Finally, the column 'Oblig. *ge*- on V3' represents the numbers of speakers who require *ge*- to appear on the lexical verb.

Type of PC	No *ge*- on V3	Opt. *ge*- on V3	Oblig. *ge*- on V3	Total
$Loop_{andative}$ en V	103 (94,5%)	6 (5,4%)	0 (0,0%)	109 (100%)
$Loop_{andative}$ V	125 (94,0%)	8 (6,0%)	0 (0,0%)	133 (100%)
Sit en V	166 (85,6%)	28 (14,4%)	0 (0,0%)	194 (100%)
Sit V	29 (80,6%)	7 (19,4%)	0 (0,0%)	36 (100%)

Table 9.4: Optionality of *ge*- on V3 for $Loop_{andative}$ and *Sit* PCs

As can be seen from the Table, the vast majority of speakers do not allow *ge*- to appear on the lexical verb, neither in the case of $loop_{andative}$, nor in the case of *sit*, regardless of whether or not *en* is present in the construction. Note, though, that the frequencies of optional *ge*- on the lexical verb are slightly higher in PC constructions with *sit* than in the ones with $loop_{andative}$. Furthermore, *ge*- never obligatorily occurs on the lexical verb, neither in the case of $loop_{andative}$, nor in the case of *sit*. This type of *ge*-lowering is thus a very infrequent and always optional phenomenon, with slightly higher frequencies for the PC verb *sit* than for $loop_{andative}$.

Let us now look at the optionality patterns of *en* per PC verb. They are given in Table 9.5. The column 'Oblig. *en*' contains the numbers of speakers who only allow PC constructions in which *en* is present. In the column 'Optional *en*' the numbers of speakers are given who allow PC constructions both with and without *en*. Finally, the column 'Oblig. no *en*' represents the numbers of speakers who only accept PC constructions in which *en* is absent. In each cas we look at the two alternatives of each PC verb, i.e. with and without *ge*-, separately, to see whether there is a possible interaction between the presence/absence of *en* and that of *ge*-. The total numbers per PC verb with or without *ge*- are calculated based on the total number of speakers for the PC verb in general (Table 9.1) minus the number of speakers for whom *ge*- is obligatory absent or obligatory present respectively (Table 9.3). So for example, for the PC verb $loop_{progressive}$ the number of speakers who accept at least one version of that test item is 161. 39 speakers require *ge-* to be present, which means that the total number of speakers that accept a $loop_{progressive}$ PC without *ge-* is 122 (161−39). On the other hand, 18 speakers need *ge-* to be absent, so the total number of speakers that accept a $geloop_{progressive}$ PC (i.e. with *ge-*) is 143 (161−18).

Verb	Oblig. *en*	Optional *en*	Oblig. no *en*	Total
$Loop_{progressive}$	83 (68,0%)	31 (25,4%)	8 (6,6%)	122 (100%)
$Ge\text{-}loop_{progressive}$	114 (79,7%)	27 (18,9%)	2 (1,4%)	143 (100%)
$Loop_{andative}$	12 (7,7%)	61 (39,1%)	83 (53,2%)	156 (100%)
$Ge\text{-}loop_{andative}$	25 (18,8%)	54 (40,6%)	54 (40,6%)	133 (100%)
Sit	149 (72,5%)	22 (27,0%)	1 (0,5%)	172 (100%)
Ge-sit	170 (87,6%)	23 (11,9%)	1 (0,5%)	194 (100%)
Staan	126 (75,0%)	37 (22,0%)	5 (3,0%)	168 (100%)
Ge-staan	128 (69,9%)	52 (28,4%)	3 (1,7%)	183 (100%)
Lê	153 (83,2%)	28 (15,2%)	3 (1,6%)	184 (100%)
Ge-lê	169 (85,4%)	28 (14,1%)	1 (0,5%)	198 (100%)

Table 9.5: Optionality of *en* per PC verb

We can observe the following patterns from the table. First, for the majority of the speakers *en* is obligatory in the case of $loop_{progressive}$ and the three posture verbs. Second, in contrast to the other PC verbs, obligatory *en* is the minority pattern in the case of $loop_{andative}$. For this verb, the dominant patterns are the obligatory absence of *en* and to a slightly lesser degree the optionality of *en*. Third, there does not seem to be an interaction between the presence/absence of *en* and the presence/absence of *ge-*: the frequencies per column are quite comparable for the versions with and without *ge-*. This means that the optionality of *ge-* and that of *en* should be seen as two quite independent phenomena.

Before moving on to the optionality patterns in the V2 PC constructions, let us briefly investigate the geographical dimension of the data just presented. Given that *ge-* is highly optional for all PC verbs and across most speakers, investigating the geographical spread of this phenomenon is unlikely to reveal any interesting patterns. By contrast, given that the absence of *en* is not equally frequent among the different PC verbs and across speakers, it is interesting to see the geographical spread of *en* absence. This is indicated on the map in Figure 9.2. On this map we see that in the middle of South Africa, only $loop_{andative}$ allows *en* to be absent. These are the red squares, situated (from left to right) on the border between the Northern Cape and North West provinces, in the lower part of the Eastern Cape province, in the lower part of the Free State province, and around Durban in the Kwazulu-Natal province. Given that the spread of *en* absence with the other PC verbs is not optimally visible on this map, let us look at two separate maps, one for the two uses of the *loop* PC, and one for the posture verbs. The former is given in Figure 9.3, and the latter in Figure 9.4.

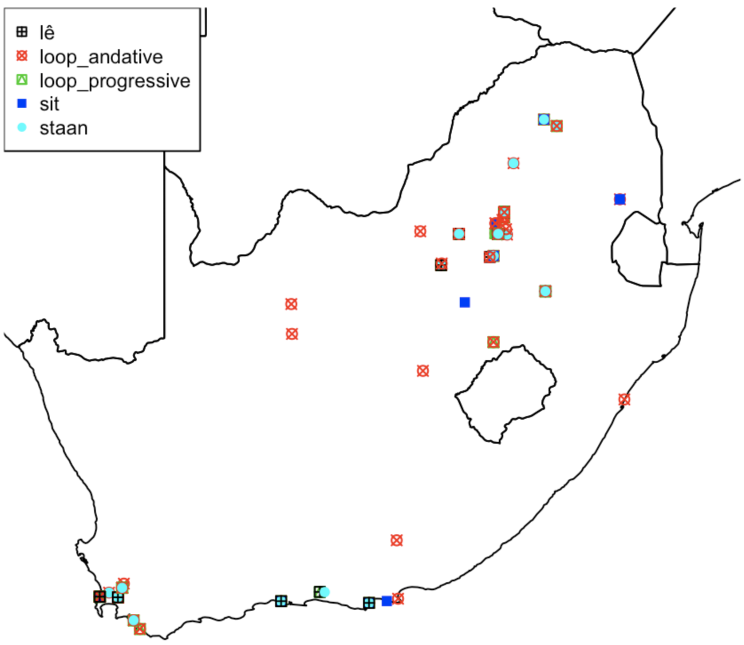


Figure 9.2: Locations in which *en* absence is allowed – all PC verbs

On the map in Figure 9.3, we see that *en*-drop is accepted with $loop_{andative}$ throughout the entire country. As for the absence of *en* with $loop_{progressive}$, this is not accepted in the middle of the country, but it is (from west to east) in and around Cape Town, on the coastal border between the Western and Eastern Cape provinces, in the northeastern part of the Free State province, in the Gauteng province, and in the central part of the Limpopo province.
Let us now look at the map in Figure 9.4, in which *en* absence is indicated for the three posture verbs. As was the case for $loop_{progressive}$, *en* absence is not accepted in the middle region of the country for the posture verbs. Furthermore, in most regions in which *en* absence is accepted with one of the posture verbs, it is also accepted with the other two. This holds (from west to east) for the region of Cape Town, the coastal boarder between the Western and Eastern Cape provinces, and the Gauteng province. Note, though, that *en* absence

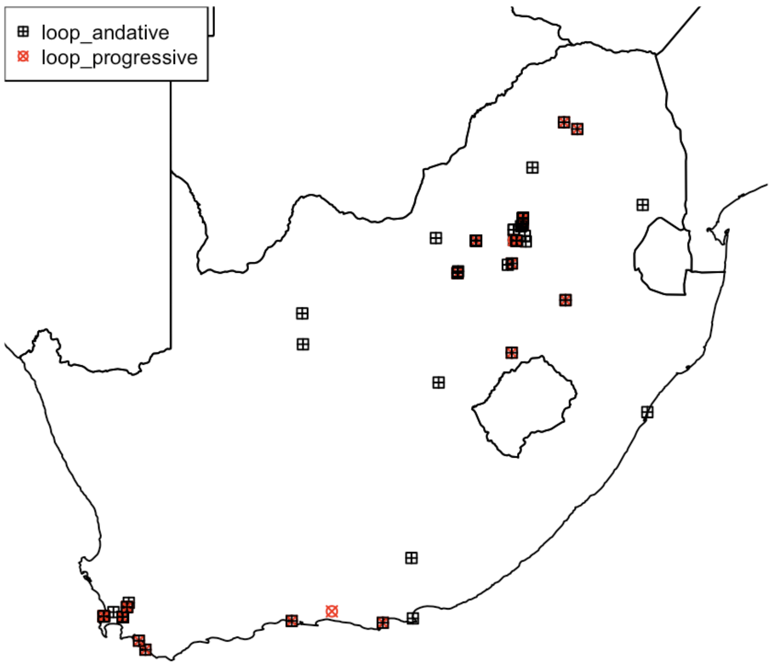


Figure 9.3: Locations in which *en* absence is allowed – *loop* PCs

with *lê* is not accepted in the Limpopo province, and that in the part of the Mpumalanga province close to the borders with Swaziland and Mozambique, *en* absence is only allowed with *sit*.

Summing up, one clear pattern emerges from the geographical spread of *en* absence in PC constructions, namely that in the middle of the country, it is only accepted with $loop_{andative}$ PCs. In the rest of the country *en* absence is also accepted with at least one, but usually more than one of the other PC verbs.

Let us now move on to the optionality patterns of the V2 PC constructions, in which either only the PC verb occurs in V2 position ('normal V2'), or the entire PC construction does ('quirky V2'). Recall that only the progressive use of *loop* was tested in these constructions, and that both configurations, normal V2 and quirky V2, were tested with and without *en* for each PC verb. The frequency distribution is given in Table 9.6. Per PC verb, the frequencies are presented separately for the constructions with and without *en*. Note, though, that there are speakers who allow both, which means that the total number of speakers per construction with and without *en* should not be seen as cu-

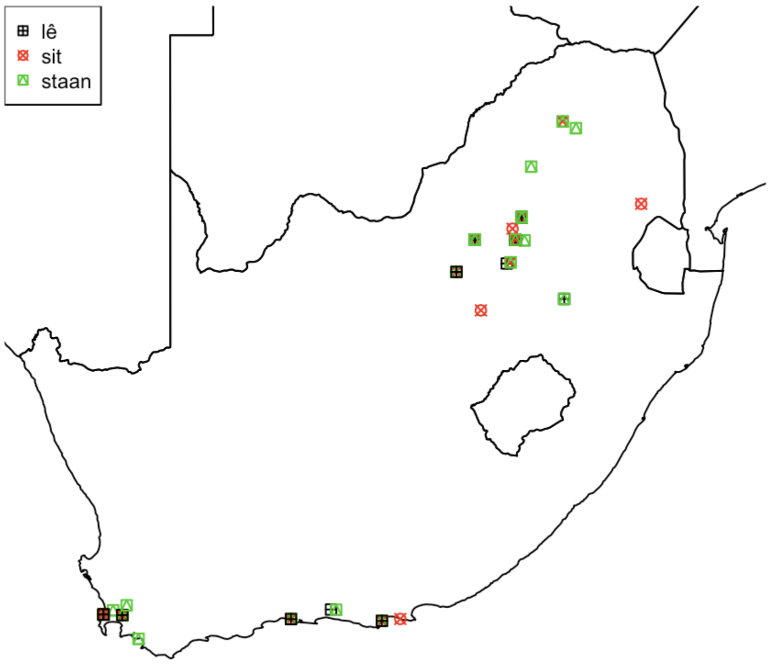


Figure 9.4: Locations in which *en* absence is allowed – posture PCs

mulative, but as partly overlapping. The column 'Oblig. normal V2' contains the number of speakers who only accept the configuration in which only the PC verb has moved to the V2 position. In the column 'Opt. normal V2' the number of speakers is given who accept both the normal and the quirky V2 configuration. Finally, the column 'Oblig. quirky V2' represents the number of speakers who only accept the quirky V2 configuration.

Let us first look at the frequencies of the PC constructions with *en*. In the table we can see that in those PC constructions, the type of V2 configuration is highly optional for all PC verbs. This is especially the case for the posture verbs, all of which have a frequency of optional normal/quirky V2 that is higher than 90%. For the PC verb $loop_{progressive}$, the frequency of optional normal/quirky V2 is lower, even though it still represents the majority of the speakers: 66,1%. For this verb, we see that 23,3% of the speakers requires a quirky V2 configuration.

Let us now focus on the configurations in which *en* is absent. First of all, we see that the total number of speakers who accept V2 configurations without *en* is much lower for all PC verbs than the number of speakers who accept V2 configurations with *en*. This is what we would expect, given that *en* absence

PC	Oblig. normal V2	Opt. normal V2	Oblig. quirky V2	Total
Loop$_{prog}$ *en* V	18 (10,2%)	117 (66,1%)	42 (23,7%)	177 (100%)
Loop$_{prog}$ V	8 (9,8%)	10 (12,2%)	64 (78,0%)	82 (100%)
Sit en V	11 (5,6%)	179 (90,9%)	7 (3,5%)	197 (100%)
Sit V	14 (36,8%)	6 (15,8%)	18 (47,6%)	38 (100%)
Staan en V	9 (4,6%)	181 (93,3%)	4 (2,1%)	194 (100%)
Staan V	10 (15,2%)	47 (71,2%)	9 (13,6%)	66 (100%)
Lê en V	6 (3,0%)	184 (92,5%)	9 (4,5%)	199 (100%)
Lê V	12 (25,0%)	12 (25,1%)	24 (50,0%)	48 (100%)

Table 9.6: Optionality of normal and quirky V2 per PC verb

was infrequent in embedded PC constructions as well (except with *loop*$_{andative}$, see Table 9.5). In Table 9.6, we can observe that V2 constructions without *en* are accepted by the highest number of speakers in the case of the PC verb *loop*$_{progressive}$ (82 speakers), followed by *staan* (66 speakers), and on the lower end of the scale by *lê* (48 speakers) and *sit* (38 speakers). In the case of *loop*$_{progressive}$, *sit* and *lê*, the quirky V2 configuration is obligatory for the majority of the speakers who accept V2 configurations without *en*. This is most clearly the case for *loop*$_{progressive}$, with 78,0% of obligatory quirky V2. As for *staan*, we see that the majority of the speakers allow both normal and quirky V2 in configurations without *en*. The general pattern that we can distill from Table 9.6 is that when *en* is present, the majority of speakers allow both normal V2 and quirky V2. For *loop*$_{progressive}$, the numbers of optional normal/quirky V2 are lower compared to the posture verbs, and *loop*$_{progressive}$ also has a group of speakers for whom quirky V2 is obligatory. When *en* is absent, the main tendency is for quirky V2 to be obligatory, even though in the case of *staan* the majority of speakers still also allows normal V2 as an option. Like with the optionality of *ge*-, given the very high frequencies of optional normal/quirky V2, it is not very interesting to investigate the geographical spread of this phenomenon, as it is unlikely to reveal any interesting patterns.

Summing up, in this subsection we have seen the following optionality patterns regarding PC constructions in Afrikaans. First, in embedded PC constructions, the presence of *ge*- is highly optional for all PC verbs. There is a slight difference between the two uses of *loop*, in the sense that the second highest frequency for *loop*$_{progressive}$ is for obligatory *ge*- presence, whereas for *loop*$_{andative}$ the second highest frequency is for obligatory *ge*- absence. Second, regarding the possible placement of *ge*- on the lexical verb rather than the PC verb – which was only tested for *loop*$_{andative}$ and *sit* – we have seen that this is a very infrequent phenomenon and never obligatory. Third, concerning the presence or absence of *en* in PC constructions, this element is by far most frequently obligatory for all PC verbs except for *loop*$_{andative}$. For this latter verb, the most frequent pattern is obligatory *en* absence, even though the frequency for optional *en* is also high. Furthermore, we have seen that there is no interaction between the presence or absence of *ge-* on the PC verb and the presence or absence of *en*. This means that these should be seen as two different phenomena. Fourth, moving on to the V2 PC constructions, normal/quirky V2 is highly optional for all PC verbs, especially for the three posture verbs. In the case of $loop_{progressive}$, normal/quirky V2 is optional for the majority of the speakers, but there is also a considerable number of speakers for whom quirky V2 is obligatory. In addition, most speakers do not accept V2 constructions in which *en* is absent, but if they do, the tendency is that quirky V2 becomes obligatory. In the next subsection, I take a closer look at the differences across PC verbs regarding these phenomena.

9.3.5 Variation and optionality across PC verbs

Let us start again with the embedded PC constructions. In this subsection, I give the weighted frequencies of the four possible configurations of embedded PC constructions and the frequency distributions over the number of options allowed. The weighted frequencies were calculated in the same way as was described in subsection 4.3.5 of case study I. I therefore do not repeat the details of the calculations here. The weighted frequencies of the four possible configurations of the embedded PC constructions are given in Table 9.7. For the sake of clarity, the four relevant configurations are repeated for $loop_{progressive}$ in (23). In the Table, the first column corresponds to the type of configuration in (23-a), the second column to (23-b), the third to (23-c), and the fourth to (23-d).[7]

(23) a. ... **loop en praat** het.
... walk and talk has
b. ... **geloop en praat** het.
... GE-walk and talk has
c. ... **loop praat** het.
... walk talk has
d. ... **geloop praat** het.
... GE-walk talk has

The following observations can be made based on Table 9.7. For all PC verbs except $loop_{andative}$, the configuration with both *ge-* and *en* has the highest weighted frequency among the four options. For these PC verbs, the second highest weighted frequency is for the configuration without *ge-*, but with *en*. The other two configurations have very comparable, low weighted frequencies. In contrast, for $loop_{andative}$, the configuration without *ge-* and without *en* has

[7] Given that *ge*-lowering, i.e. the occurrence of *ge-* on V3, was only tested for two of the five PC verbs, and furthermore turned out to be highly infrequent and always optional when possible, this configuration is not included in the frequency table.

	en present		*en* absent		
PC verb	*ge-* absent	*ge-* present	*ge-* absent	*ge-* present	Total
$Loop_{and}$	27,9 (16,8%)	28,2 (17,0%)	67,3 (40,6%)	42,6 (25,6%)	166 (100%)
$Loop_{prog}$	56,2 (34,9%)	79,4 (49,3%)	14,3 (8,9%)	11,1 (6,9%)	161 (100%)
Sit	81,1 (41,0%)	101,8 (51,4%)	7,2 (3,7%)	7,9 (3,9%)	198 (100%)
Staan	74,9 (39,0%)	87,8 (45,7%)	12,0 (6,3%)	17,3 (9,0%)	192 (100%)
Lê	83,6 (41,8%)	97,3 (48,6%)	19,1 (5,1%)	9,0 (4,5%)	200 (100%)

Table 9.7: Weighted frequencies of all configurations per PC verb - embedded sentences

the highest weighted frequency. In contrast to the other PC verbs, $loop_{andative}$ furthermore has a more equal spread among the remaining three configurations, with weighted frequencies ranging from 16,8% to 25,6%. In other words, for $loop_{andative}$ all four configurations are relatively well represented in the data, with the best represented one being 'no *ge-*, no *en*'.

Let us now look the number of options that speakers accept for each PC verb. As there are four possible configurations, the highest number of options a speaker can accept is 4. The lowest is 0, as there were some speakers per PC verb who did not accept any of the versions. The frequencies are given in Table 9.8.

PC verb	0 options	1 option	2 options	3 options	4 options	Total
$Loop_{and}$	35 (17,4%)	32 (15,8%)	59 (29,5%)	45 (22,4%)	30 (14,9%)	201 (100%)
$Loop_{prog}$	40 (19,9%)	50 (24,9%)	74 (36,8%)	24 (11,9%)	13 (6,5%)	201 (100%)
Sit	3 (1,5%)	25 (12,4%)	143 (71,1%)	20 (10,0%)	10 (5,0%)	201 (100%)
Staan	9 (4,5%)	32 (15,9%)	102 (50,8%)	30 (14,9%)	28 (13,9%)	201 (100%)
Lê	1 (0,5%)	16 (8,0%)	141 (70,1%)	31 (15,4%)	12 (6,0%)	201 (100%)

Table 9.8: Number of options per PC verb – embedded sentences

The table shows the following. For all PC verbs, two configurations is the most common option. The frequencies are lower for $loop_{andative}$ and $loop_{progressive}$ in this respect, but note that for these verbs there is also a considerable number of speakers who did not accept any version (17,4% and 19,9% respectively). There is also a difference between the two uses of *loop*. The second highest frequency for $loop_{progressive}$ (24,9%) is for one configuration, whereas it lies at 3 configurations for $loop_{andative}$ (22,4%). In addition, we see that both $loop_{andative}$ and the posture verb *staan* have relatively high frequencies for 4 configurations when compared to the other PC verbs (14,9% and 13,9% respectively). These frequencies are presented visually in the stacked bar chart in Figure 9.5. What we can see in this chart is that of the five different PC verbs, the posture verbs *sit* and *lê* behave most similarly. The posture verb *staan* behaves slightly differently from the other posture verbs in the sense that it has more speakers that accept only 1 option, and considerably more speakers that accept 4 options. The PC verb $loop_{progressive}$ is different from the others in having the highest percentage of speakers that allow only 1 option. The PC verb $loop_{andative}$ differs in having a lower number of speakers that allow 2 options compared to the other PC verbs, as well as having higher frequencies for 3 and 4 options.

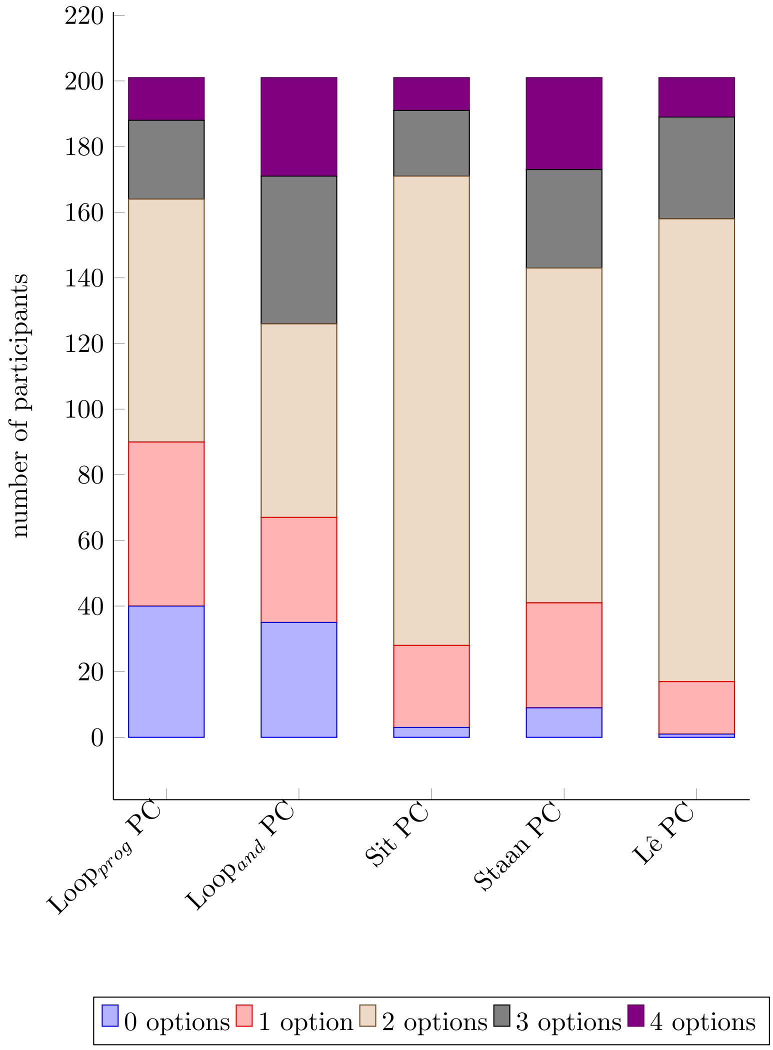


Figure 9.5: Frequencies of numbers of versions per PC verb - embedded sentences (stacked bar chart)

Let us now move on the V2 PC constructions. Recall that there were four possible configurations per PC verb, and that due to reasons of questionnaire length the PC verb $loop_{andative}$ was not tested in this condition. The four configurations are illustrated again for the PC verb $loop_{progressive}$ in (24). In Table 9.9, the first column corresponds to the type of configuration in (24-a), the second column to (24-b), the thid to (24-c), and fourth to (24-d).

(24) a. Hoekom **loop** Jan heeldag piesangs **en eet**?
Why walk.FIN Jan all.day bananas and eat
b. Hoekom **loop en eet** Jan heeldag piesangs?
Why walk.FIN and eat Jan all.day bananas
c. Hoekom **loop** Jan heeldag piesangs **eet**?
Why walk.FIN Jan all.day bananas eat
d. Hoekom **loop eet** Jan heeldag piesangs?
Why walk.FIN eat Jan all.day bananas
'Why is Jan (walking and) eating bananas all day long?'

	en present		*en* absent		
PC verb	Normal V2	Quirky V2	Normal V2	Quirky V2	Total
$Loop_{prog}$	62,1 (32,5%)	94,2 (49,4%)	6,2 (3,2%)	28,5 (14,9%)	191 (100%)
Sit	93,6 (47,3%)	90,1 (45,5%)	6,6 (3,3%)	7,7 (3,9%)	198 (100%)
Staan	87,2 (44,7%)	84,1 (43,2%)	5,5 (2,8%)	18,1 (9,3%)	195 (100%)
Lê	89,1 (44,8%)	91,5 (45,9%)	7,1 (3,6%)	11,3 (5,7%)	199 (100%)

Table 9.9: Weighted frequencies of all configurations per PC verb - V2 sentences

The following observations can be made based on this table. The configurations in the first two columns are more frequent than those in the last two for all PC verbs. This is expected as the last two columns are the configurations without *en*, which are accepted by only a small subset of the speakers. Second, for all posture verbs, the weighted frequencies for normal and quirky V2 with *en* are highly comparable. This is different for the PC verb $loop_{progressive}$, for which the quirky V2 configuration with *en* has the highest frequency. Furthermore, this verb also has a relatively high frequency of quirky V2 without *en* compared to the other PC verbs. *Staan* in turn has a higher frequency of this option compared to the other two posture verbs.

We now turn to the frequencies of the number of options per PC verb, as given in Table 9.10.

The frequency distribution shows that for all PC verbs, 2 configurations is the most frequent option. This is especially the case for *sit* and *lê* (73,1% and 68,2% respectively), and to a slightly lesser degree for *staan* (58,7%), while $loop_{progressive}$ has a considerably lower frequency (35,8%). For this last verb, there are also relatively high weighted frequencies for 1 option (29,5%) and for 3 options (25,9%). For the other three verbs, the second highest weighted frequency is 3 options, especially with *staan* (27,3% versus 14,4% with *sit* and

PC verb	0 options	1 option	2 options	3 options	4 options	Total
Loop$_{prog}$	10 (4,9%)	59 (29,5%)	72 (35,8%)	52 (25,9%)	8 (3,9%)	201 (100%)
Sit	3 (1,5%)	16 (8,0%)	147 (73,1%)	29 (14,4%)	6 (3,0%)	201 (100%)
Staan	6 (3,0%)	13 (6,5%)	118 (58,7%)	55 (27,3%)	9 (4,5%)	201 (100%)
Lê	2 (0,9%)	14 (6,9%)	137 (68,2%)	31 (18,5%)	11 (5,5%)	201 (100%)

Table 9.10: Number of options per PC verb - V2 sentences

18,5% with *lê*). Thus, in general, we see that the three posture verbs behave very similarly, though with slight differences between *staan* and the other two. *Loop*$_{progressive}$ is the most different from the other PC verbs, especially in having a relatively high weighted frequency for 1 option. These distributions are again visualised in a stacked bar chart in Figure 9.6.

As can be seen in the bar chart, *sit* and *lê* are the most similar with respect to the number of options allowed. Having 2 options is by far the most frequent pattern for these two verbs, with lower frequencies for 3 options, even lower for 1, and the lowest for 4 options. *Staan* has the same pattern, but with higher frequencies for 3 options compared to the other two posture verbs. *Loop*$_{progressive}$ differs from the posture verbs in having quite a high frequency for 1 option. It furthermore also has a relatively high frequency for 3 options, like *staan*, but unlike the other two posture verbs.

Summarising, in this subsection we have seen following variation and optionality patterns across the five PC verbs. First, regarding the embedded PC constructions, for all PC verbs except *loop*$_{andative}$, the configuration in which both *ge*- and *en* are present has the highest weighted frequency. For *loop*$_{andative}$, the highest weighted frequency goes to the configuration in which both *ge*- and *en* are absent.

A second observation we can make is that with respect to the number of configurations allowed, for all PC verbs, the highest frequency is for 2 options. However, the spread over the other numbers of options is not equal for all verbs. The PC verbs *sit* and *lê* have a very similar distribution, in which the vast majority of the speakers allow 2 options, and the other numbers of options have very low and rather equal frequencies. The third posture verb, *staan*, has a slightly different profile, as the frequencies for the other numbers of options are higher than for the other two posture verbs. For *loop*$_{progressive}$ the frequency of 2 options is lower than in the case of the posture verbs, and the second highest frequency is for 1 option. For *loop*$_{andative}$, the frequency of 2 options is even lower than that of *loop*$_{progressive}$, and moreover, the second highest frequency for *loop*$_{andative}$ is at 3 options. This verb also has a relatively high frequency for 4 options.

A third observation is that regarding the V2 PC constructions, for all posture verbs, the weighted frequencies for normal V2 with *en* and quirky V2 with *en* are highly comparable. For *loop*$_{progressive}$, quirky V2 with *en* is more fre-

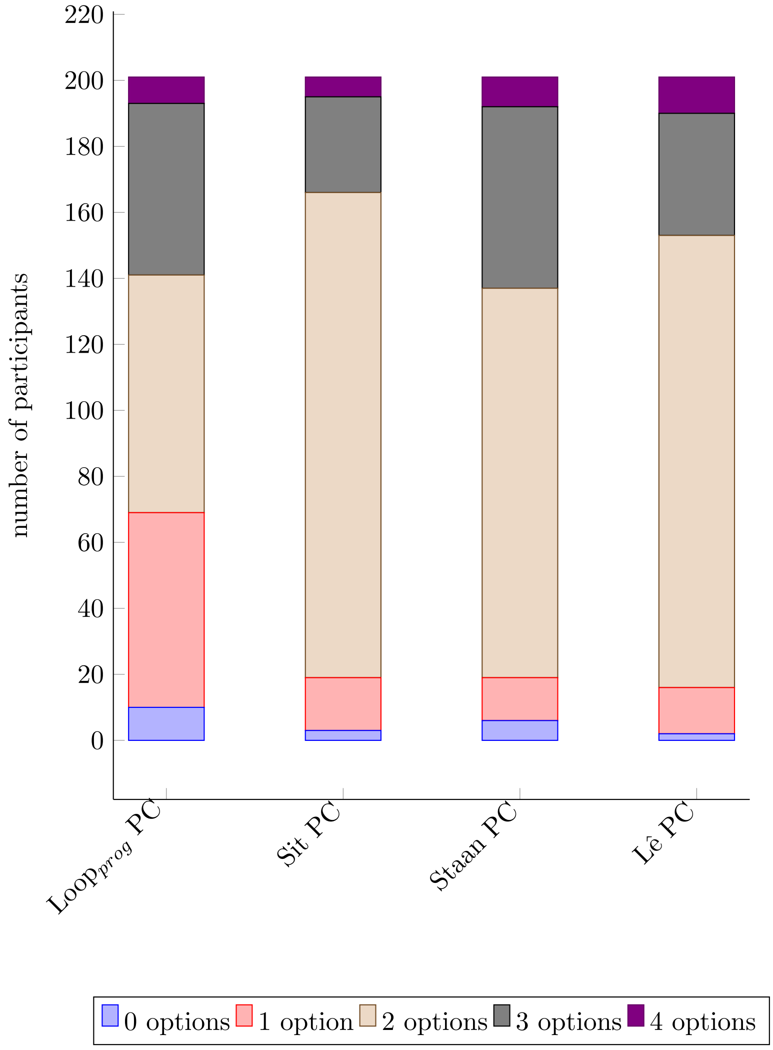


Figure 9.6: Frequencies of numbers of versions per PC verb - V2 sentences (stacked bar chart)

quent than normal V2. Furthermore, this verb has a higher weighted frequency for quirky V2 without *en* than the posture verbs.

A final observation is that in V2 constructions all PC verbs have two configurations as their most frequent option. However, the PC verb *staan* once again behaves slightly differently from the other posture verbs in having a higher frequency for 3 options. Furthermore, $loop_{progressive}$ once again differs from all posture verbs in having higher frequencies both for 1 option and for 3 options.

This verb thus shows the highest degree of variation of the four PC verbs, which was also the case in the embedded PC constructions. In those constructions, however, $loop_{andative}$ was also tested and showed an even more different profile compared to the other PC verbs.

In general we can thus conclude that the three posture verbs have a very similar profile, albeit with slightly different behavior from *staan*. $Loop_{progressive}$ behaves partly similar to the posture verbs, but also partly different, whereas $loop_{andative}$ has the most deviant profile compared of all PC verbs.

Having investigated patterns of optionality and variation in the different PC constructions in this and the previous subsection, we now move on the next subsection, in which the degree of semantic bleaching of the PC verbs is discussed. The degree to which these verbs can be combined with a semantically incompatible lexical verbs is investigated, as well as the degree to which the PC verbs still imply walking, sitting, standing or lying when they are used in a PC construction.

9.3.6 Semantic bleaching

In this subsection, I discuss the results of two sets of test items. In the first, each PC verb (except for the andative use of *loop*, see section 9.2.1) is combined with the lexical verb *swem* 'swim'. This verb is semantically incompatible with actual walking, sitting, standing or lying. For clarity's sake, the test item is repeated in (25) for the PC verb *loop*. Both a version without and a version with *ge*- were tested for each PC verb.

(25) a. Eva sê dat Pieter die laaste tyd die hele middag **loop**
Eva says that Pieter the last period the entire afternoon walk
en *swem* het.
and swim het

b. Eva sê dat Pieter die laaste tyd die hele middag
Eva says that Pieter the last period the entire afternoon
geloop en *swem* het.
GE-walk and swim het
'Eva says that lately, Pieter has been swimming entire afternoon.'

The results are given in Table 9.11.
As is clear from the table, there are very few speakers who accept the test item for the posture verbs *sit* and *lê*. In the case of $loop_{progressive}$ and *staan*, there are more such speakers, even though the large majority of them still does not accept it. The presence or absence of *ge*- does not have an effect on the acceptability of the test item.

Let us now look at the second set of test items that can inform us about the degree of semantic bleaching of the PC verbs. It concerns the additional question that was asked for all PC verbs in the V2 condition. Recall from section 9.2.1 that if speakers rated the V2 PC constructions with a 4 or a 5,

Verb	Accepted	Not accepted	Total
Loop$_{progressive}$	46 (22,9%)	155 (77,1%)	201 (100%)
Geloop$_{progressive}$	37 (18,4%)	164 (81,6%)	201 (100%)
Sit	14 (7,0%)	187 (93,0%)	201 (100%)
Gesit	17 (8,5%)	184 (91,5%)	201 (100%)
Staan	52 (25,9%)	149 (74,1%)	201 (100%)
Gestaan	52 (25,9%)	149 (74,1%)	201 (100%)
Lê	18 (9,0%)	183 (91,0%)	201 (100%)
Gelê	25 (12,4%)	176 (87,6%)	201 (100%)

Table 9.11: Number of speakers that accept incompatible lexical verb in PCs

they were asked to answer an additional question, namely whether the subject of the sentence is actually walking, sitting, standing or lying. They could answer the question with 'yes', 'no', or 'unclear'. The results are given in Tables 9.12–9.15.[8] In these tables, the column 'No xx implied' (in the case of *loop*$_{progressive}$, 'xx' stands for 'walking', for *sit* 'xx' is 'sitting' et cetera) gives the number of speakers who answered the additional question with 'no'. In the column 'xx implied' the number of speakers is given who answered the question with 'yes'. The last column gives the number of speakers for whom it was unclear whether the subject of the sentence was walking/sitting/standing/lying. The frequencies are given per type of configuration: normal V2 with *en*, quirky V2 with *en*, normal V2 without *en*, and quirky V2 without *en*. Let us look at each PC verb in turn. We start with *loop*$_{progressive}$ in Table 9.12.

Type of PC	No walking implied	Walking implied	Unclear	Total
Normal V2, *en*	52 (56,5%)	25 (27,2%)	15 (16,3%)	92 (100%)
Quirky V2, *en*	75 (59,1%)	33 (25,9%)	19 (15,0%)	127 (100%)
Normal V2, no *en*	11 (84,6%)	1 (7,7%)	1 (7,7%)	13 (100%)
Quirky V2, no *en*	50 (84,7%)	6 (10,2%)	3 (5,1%)	59 (100%)

Table 9.12: Semantic bleaching in normal and quirky V2 *loop*$_{progressive}$ PCs

The table shows that for the configurations with *en* on the one hand and those without *en* on the other, the majority option is the same. In the configurations with *en*, no actual walking is implied for more than half of the speakers. These frequencies are even higher in the configurations without *en*: more then 80% of the speakers who accept the test item think the subject of the sentence is

[8]Unfortunately, not all speakers who rated the test item with a 4 or 5 answered the additional question. This means that the total numbers per configuration are sometimes lower than the total numbers given in Table 9.6.

not actually walking. We thus see that for the majority of the speakers, the semantics of $loop_{progressive}$ are bleached in PC constructions. Let us now turn to the PC verb *sit*, for which the frequencies are given in Table 9.13.

Type of PC	No sitting implied	Sitting implied	Unclear	Total
Normal V2, *en*	34 (24,8%)	88 (64,2%)	15 (11,0%)	137 (100%)
Quirky V2, *en*	32 (24,6%)	87 (66,9%)	11 (8,5%)	130 (100%)
Normal V2, no *en*	11 (61,1%)	6 (33,3%)	1 (5,6%)	18 (100%)
Quirky V2, no *en*	4 (25,0%)	11 (68,8%)	1 (6,2%)	16 (100%)

Table 9.13: Semantic bleaching in normal and quirky V2 *sit* PCs

The table shows a different picture compared to the previous one. That is, for the configurations with *en*, the majority of the speakers who accepted the sentence think that there is actual sitting implied. Only for less than 25% of the speakers is the semantics of *sit* in the PC construction bleached to the extent that no sitting is implied. This also holds for the quirky V2 configuration without *en*, whereas for normal V2 without *en* the majority of the speakers who accept that configuration think no sitting is implied. Next, we look at the frequencies for the PC verb *staan*, given in Table 9.14.

Type of PC	No standing implied	Standing implied	Unclear	Total
Normal V2, *en*	82 (56,6%)	50 (34,5%)	13 (9,0%)	145 (100%)
Quirky V2, *en*	99 (68,8%)	31 (21,5%)	14 (9,7%)	144 (100%)
Normal V2, no *en*	13 (81,3%)	3 (8,7%)	0 (0,0%)	16 (100%)
Quirky V2, no *en*	34 (79,1%)	5 (11,6%)	4 (9,3%)	43 (100%)

Table 9.14: Semantic bleaching in normal and quirky V2 *staan* PCs

As is clear from the table, the majority of the speakers who accept the sentence in the configurations with *en* think that no standing is implied. In the configurations without *en*, these are even higher, around 80%. With respect to the semantic bleaching of *staan*, we can thus conclude that this PC verb has a profile similar to that of $loop_{progressive}$. Finally, let us look at the semantic bleaching of *lê*, given in Table 9.15.

The table shows that in all configurations, the majority of the speakers think actual lying is still implied. Note, though, that in the quirky V2 configuration without *en*, 37% of the speakers do not think lying is implied. For the other three configurations the numbers of speakers who do not think lying is implied are the highest of all PC verbs. It thus seems that *lê* is the least semantically bleached of all PC verbs, followed by *sit*, then *staan*, and finally $loop_{progressive}$ as the most bleached one. It must be noted, though, that the degree of semantic bleaching of any vocabulary item is very difficult to test, and may be different

Type of PC	No lying implied	Lying implied	Unclear	Total
Normal V2, *en*	18 (14,3%)	103 (81,7%)	5 (4,0%)	126 (100%)
Quirky V2, *en*	20 (13,8%)	122 (84,1%)	3 (2,1%)	145 (100%)
Normal V2, no *en*	3 (20,0%)	12 (80,0%)	0 (3,5%)	15 (100%)
Quirky V2, no *en*	10 (37,0%)	13 (48,1%)	4 (14,9%)	27 (100%)

Table 9.15: Semantic bleaching in normal and quirky V2 *lê* PCs

in different types of sentences and contexts. I therefore do not want to make any strong claims regarding the degree of semantic bleaching of the PC verbs based on the data presented in this subsection. On the other hand, I do think that presenting and discussing these frequencies can be interesting in light of the overall topic of this case study, namely the semi-lexical use of these verbs. In addition, the finding that the PC verbs *loop* and *staan* are more semantically bleached than *sit* and *lê* is in line with what has been reported in the recent literature on this topic, see e.g. Breed (2017) and Biberauer (2019b).

9.3.7 Statistical analysis of the patterns

9.3.7.1 Multiple Correspondence Analysis

I start out with the application of MCA. For the explanation of how this statistical technique works, the reader is referred to 4.3.6.1 of case study I. Important to clearly repeat here, though, is that the raw data table used for MCA keeps the 1-5 ratings of the Judgment task of the questionnaire. The data is thus not encoded to a binary division between 0 (= ungrammatical) and 1 (= grammatical).

The two main sets of test items, namely embedded PC constructions and V2 PC constructions, constitute the variables of the analysis. I have left out the semantic bleaching test items, as they do not provide any insight into the morphosyntactic variation in PC constructions. Recall that for the embedded PC constructions, all five PC verbs were tested in four different configurations. These configurations are (abstractly) repeated for the PC verb $loop_{progressive}$ in (26).

(26) a. …**loop en praat** het.
…walk and talk has
b. …**geloop en praat** het.
…GE-walk and talk has
c. …**loop praat** het.
…walk talk has
d. …**geloop praat** het.
…GE-walk talk has

In the MCA plots below, these four configurations are labeled as in (27) (with similar labelling for the other PC verbs).

(27) a. loop_en_V3_het_progressive
b. ge.loop_en_V3_het_progressive
c. loop_V3_het_progressive
d. ge.loop_V3_het_progressive

Given that the embedded PC constructions were tested for all five PC verbs, and that there are four different embedded PC configurations, we have (5×4) 20 variables for embedded PC constructions. In addition, for the PC verbs *loop*$_{andative}$ and *sit*, there are two additional variables, namely those in which *ge-* is positioned on the lexical verb rather than the PC verb, both with and without *en*. Recall that the second set of PC constructions, the V2 PC constructions, was not tested for the PC verb *loop*$_{andative}$. The four different V2 configurations that were tested for all other PC verbs is abstractly illustrated again for *loop*$_{progressive}$ in (28).

(28) a. **loop** . . . . . . **en eet**
walk . . . . . . and eat
b. **loop en eet** . . . . . .
walk and eat . . . . . .
c. **loop** . . . . . . **eet**
walk . . . . . . eat
d. **loop eet** . . . . . .
walk eat . . . . . .

In the MCA plots below, these four configurations are labeled as in (29), again with a similar labeling system for the other PC verbs.

(29) a. normal_V2_loop_en
b. quirky_V2_loop_en
c. normal_V2_loop_no.en
d. quirky_V2_loop_no.en

Given that the V2 PC constructions were tested for four of the PC verbs, and that this construction was tested with four different configurations, we have (4×4) 16 variables. Both sets of test items combined thus yields (20+4+16) 40 variables in the data table used for MCA. In the analysis, all variables are compared to all other variables. As a first step, again as in case study I, we first explore by means of a scree plot the percentage of variance that each dimension of the MCA explains. This plot is given in Figure 9.7.

As can be seen in the scree plot, the first dimension explains the bulk of the variance. After the second dimension, there is a sharp drop in the percentage of variance explained, followed by a steady decline from the third dimension onwards. In what follows, I therefore investigate the first two dimensions of the

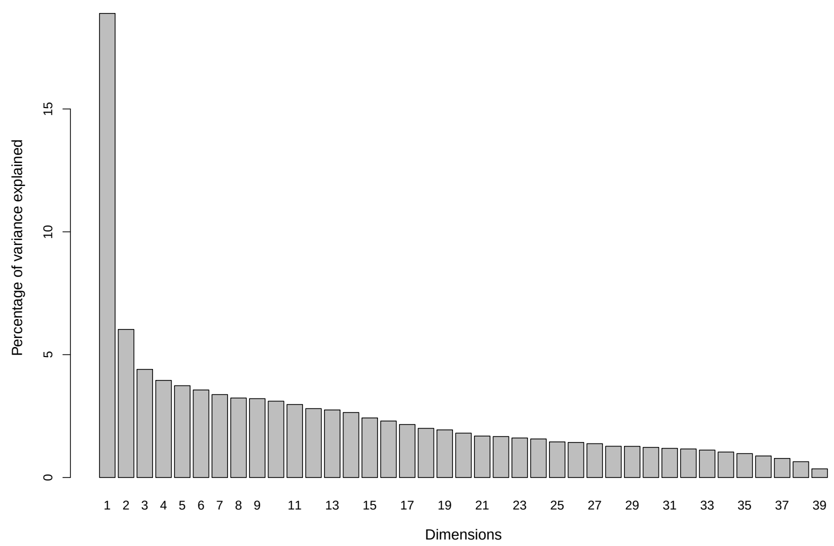


Figure 9.7: Percentage of variance explained per dimension of MCA (case study II)

MCA. The exact percentages of explained variance per dimension (of the first seven and the last dimension) are given in Table 9.16. Given that the first two dimensions are examined, this means that we get a picture of the 24.9% of the variance in the data set.

Dimension	% of variance	Cumulative % of variance
Dimension 1	18.9	18.9
Dimension 2	6.0	24.9
Dimension 3	4.4	29.3
Dimension 4	4.0	33.3
Dimension 5	3.7	37.0
Dimension 6	3.6	40.6
Dimension 7	3.4	44.0
...	...	...
Dimension 39	0.4	100

Table 9.16: Percentage of explained variance per dimension (case study II)

Given that we only examine the first two dimensions, we only have to investigate one two-dimensional plot, namely the one in which the first dimension is plotted against the second one. This plot is given in Figure 9.8.

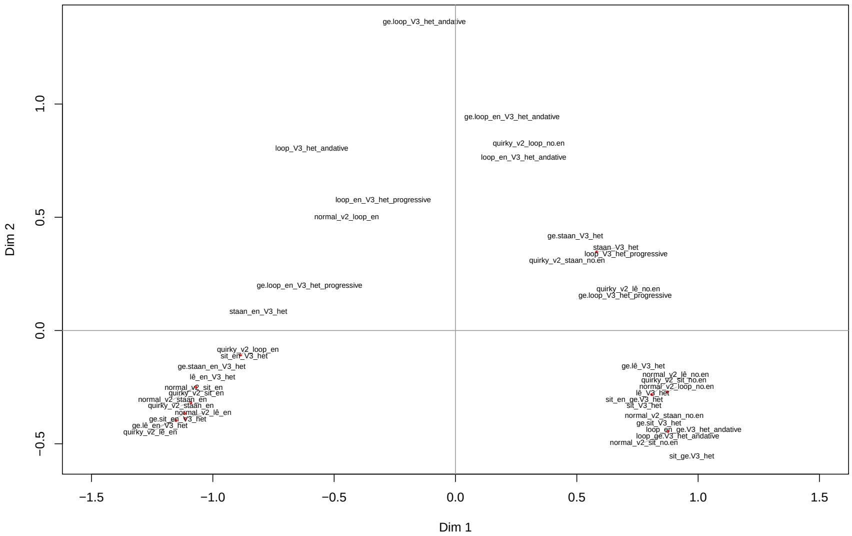


Figure 9.8: Dimension 1 and 2 of MCA (case study II)

Recall from section 4.3.6.4 in case study I that morphosyntactic configurations that are close to each other on the plot have a similar distribution across speakers, in the sense that they typically received the same rating. Configurations that are far away from each other do not have the same distribution across speakers, meaning that there is not a lot of overlap in the ratings they received. In addition, points that are located close to the origin of the plot do not contribute much to the variance in these two dimensions. Note, though, that this is irrelevant in the current plot: as one can see, there are no points that are located close to the origin. This means that all configurations contributed to the variance in the first two dimensions. In the plot, the x-axis represents the first dimension, and the y-axis the second dimension. In order to interpret the first dimension, we have to look at which variables are in the left half of the plot (to the left of the y-axis), and which ones are in the right half of the plot (to the right of the y-axis). Left of the y-axis we see the following morphosyntactic variables (from top to bottom):

1. embedded $loop_{andative}$ with *ge*-, without *en* (ge.loop_V3_het_andative)
2. embedded $loop_{andative}$ without *ge*-, without *en* (loop_V3_het_andative)
3. embedded $loop_{progressive}$ without *ge*-, with *en* (loop_en_V3_het_progressive)

4. $loop_{progressive}$ in normal V2, with *en* (normal_V2_loop_en)
5. embedded $loop_{progressive}$ with *ge-*, with *en* (ge.loop_en_V3_het_progressive)
6. embedded *staan* without *ge-*, with *en* (staan_en_V3_het)
7. $loop_{progressive}$ in quirky V2, with *en* (quirky_V2_loop_en)
8. embedded *sit* without *ge-*, with *en* (sit_en_V3_het)
9. embedded *staan* with *ge-*, with *en* (ge.staan_en_en_V3_het)
10. embedded *lê* without *ge-*, with *en* (lê_en_V3_het)
11. *staan* in normal V2, with *en* (normal_V2_staan_en)
12. *sit* in normal V2, with *en* (normal_V2_sit_en)
13. *sit* in quirky V2, with *en* (quirky_V2_sit_en)
14. *staan* in quirky V2, with *en* (quirky_V2_staan_en)
15. *lê* in quirky V2, with *en* (quirky_V2_lê_en)
16. embedded *lê* with *ge-*, with *en* (ge.lê_en_V3_het)
17. embedded *sit* with *ge-*, with *en* (ge.sit_en_V3_het)
18. *lê* in normal V2, with *en*

All these morphosyntactic configurations can be seen as 'standard' configurations, i.e. according to the literature on Afrikaans PC constructions (see chapter 7), they are accepted by all standard Afrikaans speakers: (i) all PC constructions with *en* for all PC verbs except for the andative use of *loop*, (ii) both normal and quirky V2 with *en*, and (iii) optionality of *ge-* in all the embedded configurations. In addition, we also find the *en*-less PC configurations with $loop_{andative}$, the most common option for this PC verb. These 'standard' configurations are set apart in the first dimension from the following configurations, which are positioned to the right of the y-axis in the plot (from top to bottom):

1. embedded $loop_{andative}$ with *ge-*, with *en* (ge.loop_en_V3_het_andative)
2. $loop_{progressive}$ in quirky V2, without *en* (quirky_V2_loop_no.en)
3. embedded $loop_{andative}$ without *ge-*, with *en*
4. embedded *staan* with *ge-*, without *en* (ge.staan_V3_het)
5. embedded *staan* without *ge-*, without *en* (staan_V3_het)
6. embedded $loop_{progressive}$ without *ge-*, without *en* (loop_V3_het_progressive)
7. *staan* in quirky V2, without *en* (quirky_V2_staan_no.en)
8. *lê* in quirky V2, without *en* (quirky_V2_lê_no.en)
9. embedded $loop_{progressive}$ with *ge-*, without *en* (ge.loop_V3_het_progressive)
10. embedded *lê* with *ge-*, without *en* (ge.lê_V3_het)
11. *sit* in quirky V2, without *en* (quirky_V2_sit_no.en)

12. *lê* in normal V2, without *en* (normal_V2_lê_no.en)
13. $loop_{progressive}$ in normal V2, without *en* (normal_V2_loop_no.en)
14. embedded *lê* without *ge*-, without *en* (lê_V3_het)
15. embedded *sit* with *ge*- on the lexical verb, with *en* (sit_en_ge.V3_het)
16. embedded *sit* without *ge*-, without *en* (sit_V3_het)
17. *staan* in normal V2, without *en* (normal_V2_staan_no.en)
18. embedded *sit* with *ge*-, without *en* (ge.sit_V3_het)
19. embedded $loop_{andative}$ with *ge*- on the lexical verb, with *en* (loop_en_ge.V3_het_andative)
20. embedded $loop_{andative}$ with *ge*- on the lexical verb, without *en* (loop_ge.V3_het_andative)
21. *sit* in normal V2, without *en* (normal_V2_sit_no.en)
22. embedded *sit* with *ge*-, without *en*

All these morphosyntactic configurations can be seen as 'non-standard', namely (i) *en*-absence with all PC verbs except $loop_{andative}$ en (ii) *ge*-lowering.[9] In addition, the configurations for $loop_{andative}$ with *en* are grouped together with all these other 'non-standard' configurations. The first dimension thus seems to set apart all 'standard' from all 'non-standard' configurations. Based on the first dimension, we can thus argue that *en*-absence with the andative use of *loop* really seems to be a 'standard-like' phenomenon, and that the configuration in which *en* is present, is seen as non-standard, like other non-standard phenomena such as *en*-absence with the other PC verbs, and *ge*-lowering.

Let us now examine the second dimension. To do this, we have to look at which configurations are above the x-axis (i.e. in the top part of the plot) and which ones are below (in the bottom part of the plot). The following variables are positioned above the x-axis (from left to right):

1. embedded *staan* without *ge*-, without *en* (staan_en_V3_het)
2. embedded $loop_{progressive}$ with *ge*-, with *en* (ge.loop_en_V3_het_progressive)
3. embedded $loop_{andative}$ without *ge*-, with *en* (loop_V3_het_andative)
4. $loop_{progressive}$ in normal V2, with *en* (normal_V2_loop_en)
5. embedded $loop_{progressive}$ without *ge*-, with *en* (loop_en_V3_het_progressive)
6. embedded $loop_{andative}$ with *ge*-, with *en* (ge.loop_en_V3_het_andative)

[9] Note that I describe configurations that have been mentioned in the literature on Afrikaans PC constructions as the most common configuration as 'standard', and configurations that either have not been described in the literature or that have been described as being a feature of a specific variety of Afrikaans as 'non-standard'. I am of course aware of the problems that come with labelling a certain configuration as 'standard' or 'non-standard', but use these labels in a purely descriptive sense, without wanting to claim that certain configurations are to be preferred over others from a prescriptive point of view.

7. embedded *loop*$_{andative}$ without *ge*-, with *en* (loop_en_V3_het_andative)
8. *loop*$_{progressive}$ in quirky V2, without *en* (quirky_V2_loop_no.en)
9. *staan* in quirky V2, without *en* (quirky_V2_staan_no.en)
10. embedded *staan* with *ge*-, without *en* (ge.staan_V3_het)
11. embedded *staan* without *ge*-, without *en* (staan_V3_het)
12. embedded *loop*$_{progressive}$ without *ge*-, without *en* (loop_V3_het_progressive)
13. embedded *loop*$_{progressive}$ with *ge*-, without *en* (ge.loop_V3_het_progressive)
14. *lê* in quirky V2, without *en* (quirky_V2_lê_no.en)

This list contains all morphosyntactic configurations that involve the PC verbs *loop*$_{andative}$ and *loop*$_{progressive}$, with three exceptions: (i) *ge*-lowering with *loop*$_{progressive}$, (ii) normal V2 without *en* with *loop*$_{andative}$, and (iii) quirky V2 with *en* with *loop*$_{andative}$. In addition, there are four configurations of the PC verb *staan* on this side of the x-axis, three of which are configurations without *en*. This shows that of the posture PC verbs those with *staan* behave most similarly to the ones with *loop*, which confirms our earlier findings in subsections 9.3.4–9.3.6. The odd one out in the list in (29) is the quirky V2 configuration for *lê* in which *en* is absent. However, it is very closely positioned to that same configuration with *staan*. In general, though, this group of configurations are all the *loop*$_{andative}$ and *loop*$_{progressive}$ ones that are most widely accepted (with the exception of quirky V2 with *en* for *loop*$_{progressive}$, which is situated below the x-axis). In addition to this, we find mostly non-standard configurations with *staan*, and exceptionally one non-standard *lê* configuration. These configurations are set apart in the second dimension from the ones below the x-axis, which are (from left to right):

1. *lê* in normal V2, with *en* (normal_V2_lê_en)
2. embedded *sit* with *ge*-, with *en* (ge.sit_en_V3_het)
3. embedded *lê* with *ge*-, with *en* (ge.lê_en_V3_het)
4. *lê* in quirky V2, with *en* (quirky_V2_lê_en)
5. *staan* in quirky V2, with *en* (quirky_V2_staan_en)
6. *sit* in quirky V2, with *en* (quirky_V2_sit_en)
7. *sit* in normal V2, with *en* (normal_V2_sit_en)
8. *staan* in normal V2, with *en* (normal_V2_staan_en)
9. embedded *lê* without *ge*-, with *en* (lê_en_V3_het)
10. embedded *staan* with *ge*-, with *en* (ge.staan_en_V3_het)
11. embedded *sit* without *ge*-, with *en* (sit_en_V3_het)
12. *loop*$_{progressive}$ in quirky V2, with *en* (quirky_V2_loop_en)
13. embedded *sit* with *ge*- on lexical verb, with *en* (sit_en_ge.V3_het)

14. embedded *sit* without *ge*-, without *en* (sit_V3_het)
15. embedded *lê* with *ge*-, without *en* (ge.lê_V3_het)
16. embedded *lê* without *ge*-, without *en* (lê_V3_het)
17. *sit* in quirky V2, without *en* (quirky_V2_sit_no.en)
18. *lê* in normal V2, without *en* (normal_V2_lê_no.en)
19. $loop_{progressive}$ in normal V2, without *en* (normal_V2_loop_no.en)
20. *staan* in normal V2, without *en* (normal_V2_staan_no.en)
21. embedded *sit* with *ge*-, without *en* (ge.sit_V3_het)
22. *sit* in normal V2, without *en* (normal_V2_sit_no.en)
23. embedded $loop_{andative}$ with *ge*- on lexical verb, without *en* (loop_ge.V3_het_andative)
24. embedded $loop_{andative}$ with *ge*- on lexical verb, with *en* (loop_ge.V3_en_het_andative)
25. embedded *sit* with *ge*-, without *en* (sit_ge.V3_het)

These are all configurations with posture verb PCs, expect for the ones mentioned above: three non-standard *staan* configurations, one standard *staan* configuration, and one non-standard *lê* configuration. In addition, there are a few exceptions of configurations with both uses of *loop*. In general, though, this list contains all configurations for the PC verbs *sit* and *lê*, and the majority of the configurations for the PC verb *staan*. The second dimension thus sets apart both uses of the PC verb *loop* from the posture verbs, with the posture verb *staan* being the least different from the two *loop*'s compared to the other two posture verbs.

Before we dive deeper into the clusters of configurations in the next subsection, it is worth noting here that neither in the first nor in the second dimension are configurations with and without *ge*- set apart from one other. This is different from the presence of absence of *en*, which plays an important role in the first dimension. This thus means that *ge*-presence, as we have already seen in the previous subsections, is truly optional, and a separate phenomenon from *en* presence or absence.

9.3.7.2 Hierarchical clustering: configurations

In this subsection, I use Hierarchical Clustering (HC) on the output of the MCA to investigate the possible subgroups of configurations. For a brief explanation of how HC works, the reader is again referred back to 4.3.6.2.[10] When we use HC on the output of the current MCA, the algorithm proposes a cut off point at six clusters of configurations, as can be seen by the black thick line in the hierarchical tree in Figure 9.9.

[10] The calculations of this subsection were carried out in R (R Core Team 2014) using the HCPC function of the FactoMineR package (Husson et al. 2014).

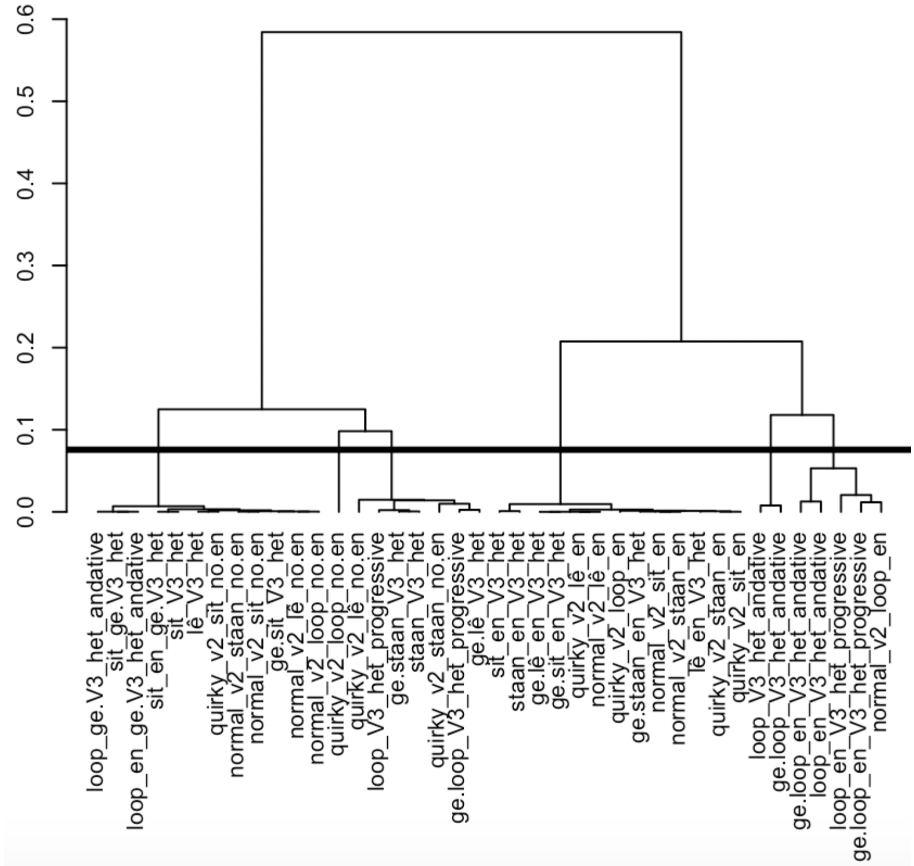


Figure 9.9: Hierarchical tree based on MCA output (case study II)

Given that in this case study 40 configurations were included in the MCA, and thus also in the HC, six clusters seems like a reasonable cut off point. That is, there will most likely be no clusters that consist of just one configuration. Let us now examine which configurations are clustered together if we use six clusters. The clusters are given in Figure 9.10.

As can be seen in the plot, there are nonetheless two very small clusters: cluster 2 (the red one) which consists of two configurations (the two embedded $loop_{andative}$ configurations without *en*), and cluster 4 (dark blue) consisting of one configuration (quirky V2 $loop_{progressive}$ without *en*). The other four clusters are bigger. The first cluster (black) consists of all embedded posture verb configurations with *en*, and all normal and quirky V2 configurations with *en*, except for the normal V2 configuration of $loop_{progressive}$. The third cluster (green) consists of normal V2 of $loop_{progressive}$ with *en*, both embedded config-

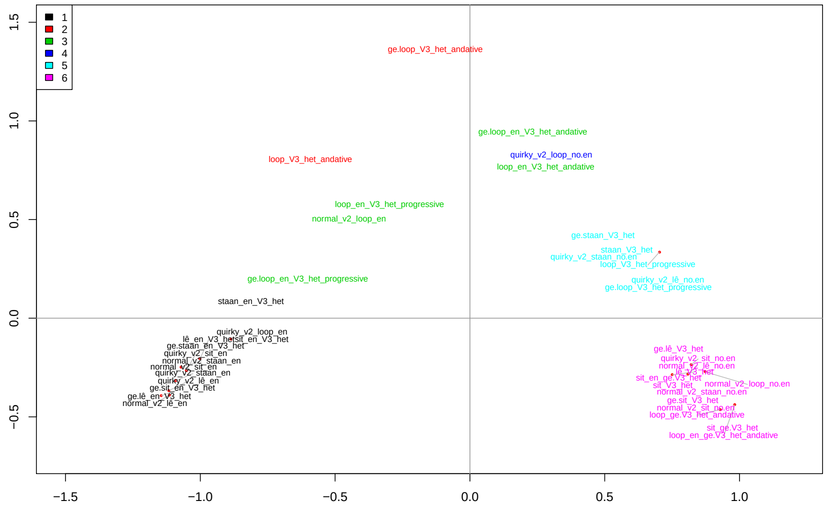


Figure 9.10: Configurations clustered into six groups (case study II)

urations of $loop_{progressive}$ with *en*, and the same configurations for $loop_{andative}$. The fifth cluster (light blue) consists of the embedded configurations for *staan* and $loop_{progressive}$ without *en*, as well as quirky V2 without *en* for *staan* and *lê*. The sixth cluster (pink) consists of all embedded configurations of *sit* and *lê* without *en*, all configurations of *ge*-lowering, and some of the normal and quirky V2 configurations without *en* for $loop_{progressive}$, *sit*, *staan* and *lê*. The clusters can generally be described as follows: cluster 1 contains all 'standard' configurations for the posture verbs, cluster 2 the 'standard' configurations for $loop_{andative}$, cluster 3 both 'standard' and 'non-standard' configurations for both uses of *loop*, cluster 4 the Quirky V2 configuration of $loop_{progressive}$ without *en*, cluster 5 the 'non-standard' configurations of $loop_{progressive}$ and *staan*, and cluster 6 all other 'non-standard' configurations.

Even though this clustering into six clusters is already quite informative, let us see whether the configurations in cluster 2 ('standard' $loop_{andative}$ configurations) and cluster 4 (the well accepted quirky V2 configuration for $loop_{progressive}$ without *en*) remain two separate clusters at a lower cut off point, or whether they are joined into a single cluster. Let us start by placing the cut off point at five clusters. The relevant plot is given in Figure 9.11.

What we can observe from this plot is that the singleton cluster that we saw in Figure 9.10, the quirky V2 configuration of $loop_{progressive}$ without *en*, is now put together with cluster 4 (the dark blue cluster). This cluster contains

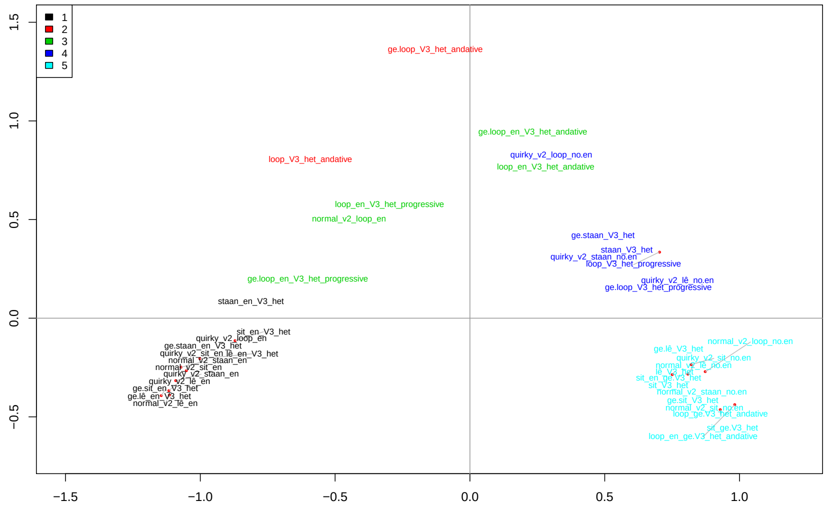


Figure 9.11: Configurations clustered into five groups (case study II)

'non-standard' configurations with *loop*$_{progressive}$ and *staan*. This means that the quirky V2 configuration of *loop*$_{progressive}$ without *en* is grouped together with other 'non-standard' configurations of this PC verb at this more coarse grained clustering. All other clusters still contain the same configurations as in the previous plot. This means that those four clusters (cluster 1, 2, 3 and 5 in this plot) are quite stable. We can investigate which of them is most stable by looking at an even more coarse grained level of clustering, namely by looking at the distribution of configurations over four clusters. The plot for this clustering is given in Table 9.12.

In the plot, we can see that the two clusters that are still the same compared to the previous two are the cluster with only 'non-standard' configurations of all PC verbs (here cluster 4, in the previous plot cluster 5), and the cluster of 'non-standard' configurations of *staan* and *loop*$_{progressive}$ (here cluster 3, in the previous plot cluster 4). Given that this cluster was only formed at the cut off point of five clusters, and did not exist at the cut off point of six, the most stable cluster so far is cluster 4 in this plot. This means that the most stable cluster is the one containing all types of 'non-standard' configurations for all PC verbs. For the other clusters in the previous plot, we see that they have all changed. Cluster 1 has undergone the fewest changes: it now contains one extra configuration, namely embedded *loop*$_{progressive}$ with *ge-* and *en*. Cluster 2 in the current plot combines two clusters of the previous plot, the one with the

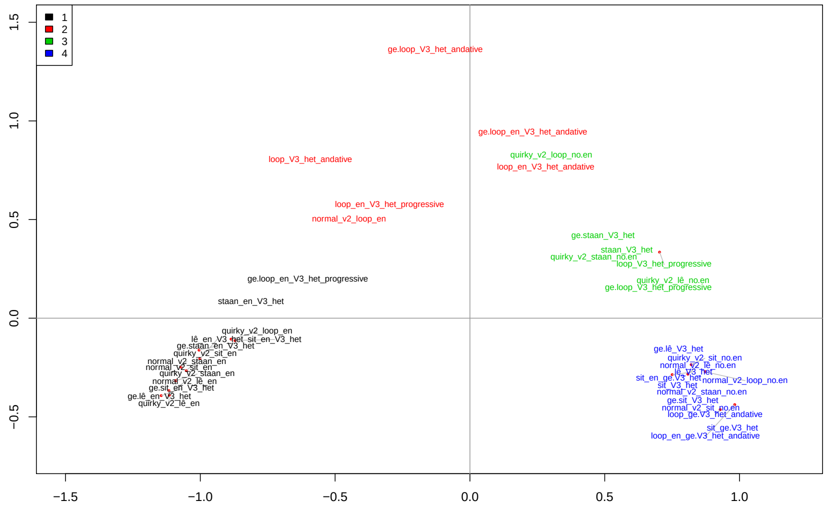


Figure 9.12: Configurations clustered into four groups (case study II)

'standard' configurations of $loop_{andative}$ and the one with 'standard' and 'non-standard' configurations of both uses of *loop*. Thus, at this level of clustering, all configurations of $loop_{andative}$ are taken together with 'standard' configurations of $loop_{progressive}$.

Let us go even one level lower, and set the cut off point at three clusters, to see whether the most robust cluster so far – the one containing many but not all 'non-standard' configurations for all PC verbs – remains a separate cluster. The plot with 3 clusters is given in Figure 9.13.

In the plot, we can see that at this coarse level of clustering, cluster 1 remains the same. As for cluster 2, the quirky V2 configuration of $loop_{progressive}$ without *en* gets added to it, making this a cluster of both 'standard' and 'non-standard' configurations of both uses of *loop*. Cluster 3 now comprises all types of 'non-standard' phenomena, including even more 'non-standard' configurations of PC verbs *staan* and $loop_{progressive}$, and one 'non-standard' configuration of *lê*. This means that at the coarse granulariy level of three clusters, cluster 1 contains all 'standard' configurations of all posture verbs, and a number of standard configurations of $loop_{progressive}$, cluster 2 contains both 'standard' and 'non-standard' configurations of both uses of *loop*, and cluster 3 contains almost all 'non-standard' configurations for $loop_{progressive}$, and all three posture verbs.

Given that at five clusters, there is not a single one that contains only one configuration, let us take this level as our baseline. Note that of these five

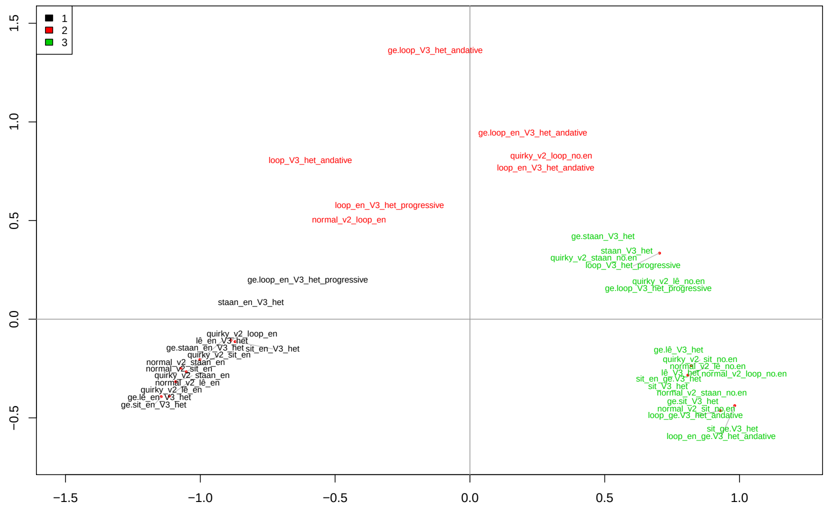


Figure 9.13: Configurations clustered into three groups (case study II)

clusters, the fifth one is the most robust: it remains the same at the the more coarse-grained level of four clusters as well. The five clusters can be described as follows:

1. all 'standard' configurations of the posture verbs, combined with the 'standard' quirky V2 configuration of $loop_{progressive}$ with *en*

2. all 'standard' configurations of $loop_{andative}$

3. 'standard' configurations of $loop_{progressive}$ and 'non-standard' configurations of $loop_{andative}$

4. 'non-standard' configurations of $loop_{progressive}$ and *staan*, in combination with one 'non-standard' configuration of *lê*

5. 'non-standard' configurations, in particular *en*-absence and *ge*-lowering, for mostly *sit* and *lê*, but also some of the other PC verbs

Let us now turn to the next subsection, where we investigate whether there are groups of speakers that can be distinguished.

9.3.7.3 Hierarchical clustering: speakers

In this subsection, I briefly investigate the geographical distribution of certain groups of speakers. In order to do so, I have applied HC on a MCA that uses the speakers rather than the configurations as individuals (again, see 4.3.6.3 in case study I for a more detailed explanation). The algorithm suggests to set the cut off point at five clusters. If we do, we get the plot in 9.14.

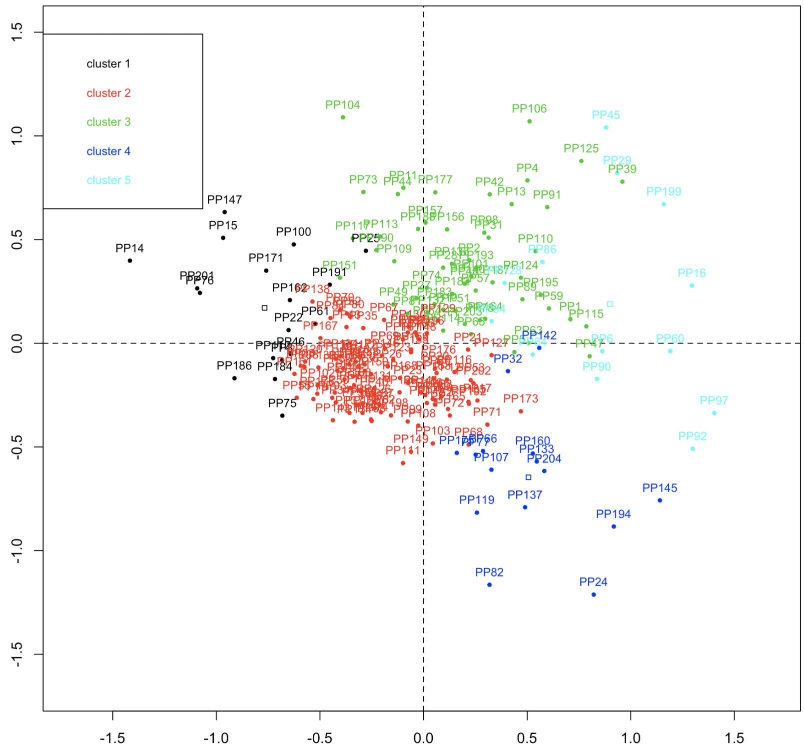


Figure 9.14: Speakers clustered into five groups (case study II)

Recall that in (M)CA plots, the data points that are closest to the origin are those that contributed least to the variance in the data. This means that the red cluster of speakers can be seen as those who have the most uniform behaviour compared to the other groups of speakers. In the plot, we can furthermore see that this cluster is by far the largest one. After that, cluster 3 (the green cluster) is also quite large. The other three clusters contain much fewer speakers.

In case study I, I have investigated the paragons of the clusters of speakers that came out of the HC. Paragons are the most representative individuals of a cluster. However, given that in this case study 40 configurations were included in the analysis, it would not be very meaningful to list the configurations allowed by the paragons of these five clusters, as it would be too much information to process. Given that the red cluster is the most uniform one, and adds the least to the variance in the data set, what we can do instead is investigate the geographical spread of this cluster compared to the other ones. This spread is captured on the map in Figure 9.15.

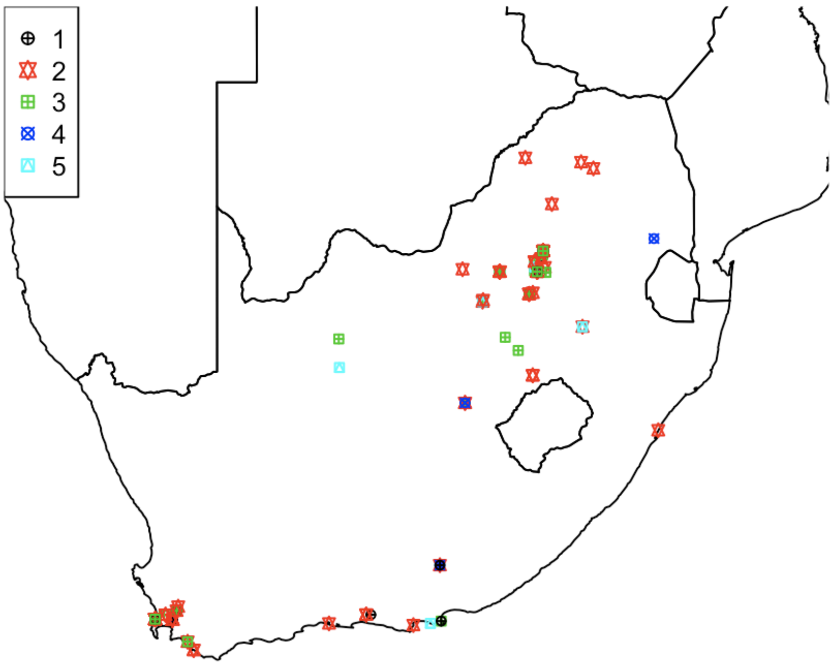


Figure 9.15: Geographical distributions of the five groups of speakers (case study II)

On the map we can see that the red dots of cluster 2 are positioned in all locations except from the middle of the country and in one location in the Mpumalanga province. These places thus have a more varied profile compared to the rest of the country.

9.3.7.4 Correspondence analysis

In this subsection, I apply CA to the data of the questionnaire study. Recall from 4.3.6.1 of case study I that CA requires the data to be binary encoded. By using a binary opposition between accepted and not-accepted, the distance (and thus (dis)similarity) between (groups of) configurations can be seen more sharply than by adhering to the original 1-to-5-ratings, as was done in the execution of MCA. Thus, for the execution of CA, I have used the binary encoded data set, in which the ratings of 1, 2 and 3 were encoded as 0, and the ratings of 4 and 5 as 1. For details on the exact steps taken in CA, see 4.3.6.1.

We start by examining the scree plot of the CA. It is given in Figure 9.16.

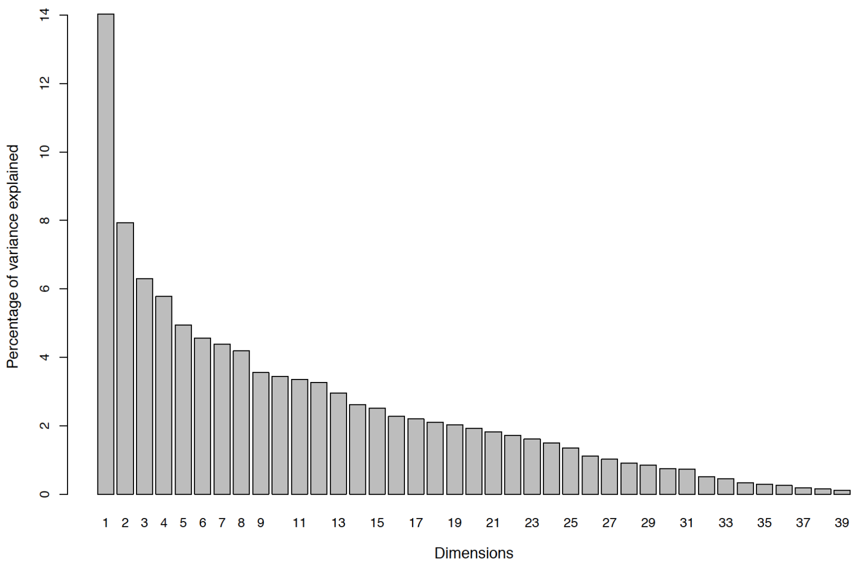


Figure 9.16: Percentage of variance explained per dimension of CA (case study II)

The scree plot shows that the first dimension explains the bulk of the variance. After the second dimension, there is a sharp drop in the percentage of variance explained, followed by a more or less steady decline from the third dimension onwards. I therefore consider only the first two dimensions.

Let us examine two-dimensional plot of the first and second dimension of the CA. It is given in 9.17. The x-axis represents the first dimension, and the y-axis the second. In the plot, we see that there is a large number of configurations which are located quite close to the origin of the plot, which means that they did not contribute a lot to the variance in the data. Examining the configurations which are grouped together in this way, we can see that they are all the configurations which are 'standard' language, and are generally accepted by all speakers. They include all quirky and normal V2 configurations with *en* and all embedded PC constructions with *en*, both with and without *ge-*. Note that for $loop_{andative}$, both embedded constructions with and without *en* are located close to the origin. Recall that in order to interpret the first dimension, we have to look at which configurations are to the left of the y-axis, and which are to the right of this line. This means that the first dimension sets apart all 'standard' V2 and embedded PC constructions, including $loop_{andative}$ PC constructions with *en*, and *ge*-lowering with $loop_{andative}$ without *en*, from all configurations of $loop_{progressive}$, and the three posture verbs without *en*, *ge*-lowering with $loop_{andative}$ with *en*, and both *ge*-lowering configurations with *sit*. In other words, the first dimension separates the 'standard' from the 'non-standard' configurations, a few exceptions notwithstanding.

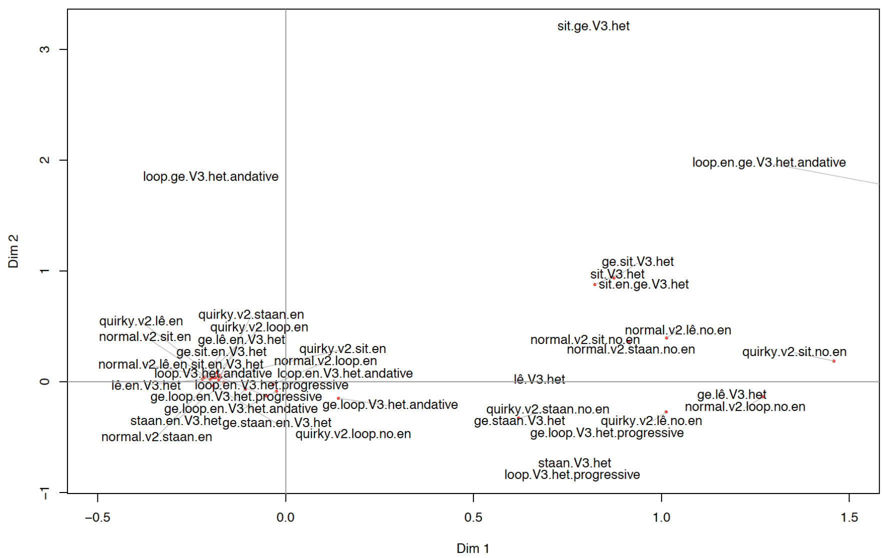


Figure 9.17: Dimension 1 and 2 of CA (case study II)

Let us now examine the second dimension by looking at the configurations above and below the x-axis. All the *ge*-lowering configurations (both for $loop_{andative}$ and for *sit*) are above the x-axis, and are thus set apart from other 'non-standard' configurations. Besides the configurations with *ge*-lowering, we also find all 'non-standard' configurations of the PC verb *sit*, and the 'non-standard' normal V2 configurations without *en* of *lê* and *staan* above the x-axis. In the bottom half of the plot, we see all 'non-standard' configurations for $loop_{progressive}$, and some 'non-standard' configurations for *lê* and *staan*. In other words, the second dimension shows that in the 'non-standard' configurations, $loop_{progressive}$ and *sit* are the most distinct from each other, with *staan* and *lê* occupying in between positions. This is in line with the rest of this chapter, in which we have seen that of the four progressive verbs, *loop* behaves the most differently from *sit*, and the two other posture verbs take up intermediate positions.

Summing up, based on the first two dimensions of the CA, we have seen (i) that all 'standard' configurations are grouped together, including PC constructions with and without *en* for $loop_{andative}$, (ii) that the 'non-standard' configurations with *ge-* on the lexical verb are partly separated from the other 'non-standard' configurations, and (iii) that in the 'non-standard' group of configurations, the ones of $loop_{progressive}$ and *sit* are most distant from each other, while the ones of *staan* and *lê* are positioned in between the two extremes.

9.3.8 Results of corpus and questionnaire study compared

In this subsection, I briefly compare the results of the corpus study and the questionnaire study. Recall that in the corpus study, only embedded PC constructions were extracted, which means that I will only compare these with the embedded PC constructions that were tested in the questionnaire. The main findings of the corpus study were as follows:

1. *Loop* behaves significantly differently from the posture verbs
2. *Ge-* is completely optional in PC constructions with posture verbs
3. *En* is frequently absent in PC constructions with *loop*, and never with the posture verbs
4. The type of aspect expressed by *loop* has a clear effect on the morphosyntactic configuration of the PC construction:
    - When expressing andative aspect, the most frequent configuration for *loop* is without *ge-* and without *en* (i.e. '*loop* V')
    - When expressing progressive aspect, two configurations have equally high frequencies, namely the one without *ge-* and without *en* (i.e. '*loop* V') and the one with *ge-* and with *en* (i.e. '*geloop en* V')

The main findings of the questionnaire study can be summarized as follows.

1. Of all PC verbs, $loop_{andative}$ behaves the most differently from all others, with $loop_{progressive}$ occupying an intermediate position between $loop_{andative}$ and the posture verbs
2. All posture verbs behave very similarly, albeit that *staan* is slightly closer to $loop_{progressive}$ then the other posture verbs
3. In general, *ge-* is highly optional for all PC verbs. *Ge-* has the highest relative frequencies for $loop_{progressive}$, the lowest for $loop_{andative}$, and frequencies of around 50% for the posture verbs

4. *En* is most frequently obligatorily absent with $loop_{andative}$, and most frequently obligatorily present with all other PC verbs, including $loop_{progressive}$

If we compare the main findings of the corpus study to those of the questionnaire study, we see that they are generally the same. A small difference is that in the corpus study, the configuration without *ge-* and without *en* was also found to be frequent for the progressive use of *loop*, whereas in the questionnaire study, this configuration did not have high frequencies. This difference can be due to the fact that in the corpus study, the type of aspect had to be annotated by two native speakers, which is a far less direct method of controlling for the type of aspect than by creating specific test items for each type. A second small difference between the results of the corpus study and questionnaire is that in the former, there were zero hits for the absence of *en* with the posture verbs, whereas in the latter, there were low frequencies of *en* absence with these verbs. This difference might be due to two factors. First, the corpus data might not include text or speech from specific locations in which *en* absence with posture verbs is allowed. Second, corpus data generally contain fewer options than questionnaire data, because in a questionnaire, a speaker can rate multiple versions of a sentence as acceptable, whereas a corpus typically does not contain (many instances of) less preferred options. Note, however, that the weighted frequencies of *en* absence with posture verbs was very low, which also reflects that these configurations are generally dispreferred. Summing up, the results of the corpus study support the results of the questionnaire study, and vice versa.

9.3.9 Patterns to be explained in the analysis

In this subsection, I present the patterns of morphosyntactic variation and optionality found in the corpus and questionnaire study that should be explained by the formal analysis. In order to do so, I first set a threshold for which frequencies in the results of the questionnaire study I take to be noise. As I have done in case study I, I draw the line at a weighted frequency of 5%. That is, all configurations that have a weighted frequency lower than 5% are taken to be noise in the questionnaire data, and do not need to be explained by the formal analysis. Let us therefore briefly return to the tables of the weighted frequencies, of the embedded PC constructions and the V2 PC constructions. The former is repeated here in Table 9.17, the latter in Table 9.18.

The threshold being set at 5%, the following embedded PC configurations are taken to be noise from now on: (i) both configurations without *en* (with and without *ge-*) for the PC verb *sit* and (ii) the configuration with *ge-* and without *en* for the PC verb *lê*. Note that the configuration without *ge-* and without *en* for this latter verb is just above the threshold, with a weighted frequency of 5,1%.

Doing the same for the V2 PC constructions, the following V2 configurations are taken to be noise from now on: (i) the normal V2 configuration without *en*

PC verb	*en* present		*en* absent		Total
	ge- absent	*ge-* present	*ge-* absent	*ge-* present	
$Loop_{and}$	27,9 (16,8%)	28,2 (17,0%)	67,3 (40,6%)	42,6 (25,6%)	166 (100%)
$Loop_{prog}$	56,2 (34,9%)	79,4 (49,3%)	14,3 (8,9%)	11,1 (6,9%)	161 (100%)
Sit	81,1 (41,0%)	101,8 (51,4%)	7,2 (3,7%)	7,9 (3,9%)	198 (100%)
Staan	74,9 (39,0%)	87,8 (45,7%)	12,0 (6,3%)	17,3 (9,0%)	192 (100%)
Lê	83,6 (41,8%)	97,3 (48,6%)	19,1 (5,1%)	9,0 (4,5%)	200 (100%)

Table 9.17: Weighted frequencies of all configurations per PC verb - embedded sentences

PC verb	*en* present		*en* absent		Total
	Normal V2	Quirky V2	Normal V2	Quirky V2	
$Loop_{prog}$	62,1 (32,5%)	94,2 (49,4%)	6,2 (3,2%)	28,5 (14,9%)	191 (100%)
Sit	93,6 (47,3%)	90,1 (45,5%)	6,6 (3,3%)	7,7 (3,9%)	198 (100%)
Staan	87,2 (44,7%)	84,1 (43,2%)	5,5 (2,8%)	18,1 (9,3%)	195 (100%)
Lê	89,1 (44,8%)	91,5 (45,9%)	7,1 (3,6%)	11,3 (5,7%)	199 (100%)

Table 9.18: Weighted frequencies of all configurations per PC verb - V2 sentences

for all PC verbs, and (ii) the quirky V2 configuration without *en* for the PC verbs *sit* and *lê*.

Having separated the signal from the noise in the data, we arrive at the following patterns of morphosyntactic variation and optionality in PC constructions that should be explained by the formal analysis:

1. Of all PC verbs, $loop_{andative}$ has the most distinct morphosyntactic behaviour. $Loop_{progressive}$ occupies an intermediate position between $loop_{andative}$ and and the posture verbs, and of the posture verbs, the behaviour of *sit* and *lê* is the most similar. *Staan* is slightly closer in behaviour to the $loop_{progressive}$. In other words, we have the following morphosyntactic scale:[11]

   $loop_{andative}$ > $loop_{progressive}$ > *staan* > *lê*/*sit*

2. $Loop_{andative}$ has the highest degree of variation and optionality in PC constructions, followed by $loop_{progressive}$ and *staan*, and with *sit* and *lê* showing the least amount of variation

[11] Note that the order in this scale is exactly the same as the one proposed for Afrikaans PC verbs in (Biberauer 2019b:12). A more coarse-grained scale is proposed in (De Vos 2001:53). He makes no distinction between *staan* and the other two posture verbs, but other than that the scale he proposes is identical to the one given here.

3. Regarding the presence/absence of *en*:
   - *En* is never absent with *sit*, and almost never with $loop_{progressive}$, *staan* and *lê*
   - *En* is very frequently absent with $loop_{andative}$, even though it can also be present
4. Regarding the presence/absence of *ge*-:
   - *Ge*- is highly optional for all posture verbs
   - *Ge*-presence is the most frequent option for $loop_{progressive}$, but *ge*-absence is also frequent for this verb
   - *Ge*-absence is the most frequent option for $loop_{andative}$, but *ge*-presence is also frequent for this verb
   - The presence/absence of *ge*- does not interact with that of *en*, and should thus be taken as an independent phenomenon
   - *Ge*- is optional both in 'standard' and in 'non-standard' Afrikaans
5. Regarding Normal versus Quirky V2:
   - For all posture verbs both configurations are highly optional
   - For $loop_{progressive}$ the frequency of Quirky V2 is higher than that of Normal V2

With these five main findings in place, I move on to the next chapter, in which I present the formal analysis of these findings.

CHAPTER 10

Analysis

10.1 Introduction

In this chapter, I present the formal analysis of this case study. In section 10.2, I briefly provide more background information about Afrikaans functional and semi-lexically used verbs. In section 10.3, I establish the stage of semi-lexicality the PC verbs are in. Sections 10.4–10.6 are devoted to the morphosyntactic status of three elements that are part of (some or all) PC constructions that are investigated in this case study: (i) the clause-final auxiliary *het* 'have', (ii) the perfect participle marker *ge-*, and (iii) the element that combines the PC verb with the lexical verb, namely *en*. I propose that the first two elements are both past tense markers in modern day Afrikaans. For *en*, I propose that it is a linker element, which is part of the PC construction, but does not project. In section 10.7, I present the syntactic structures I propose for PC configurations. In section 10.8, I illustrate how these proposed structures can explain the patterns found in this case study.

10.2 Afrikaans functional and semi-lexically used verbs

Before diving into the underlying syntactic structures of the PC constructions, I first need to give a brief overview of the Afrikaans verbal system more generally. That is, it is important to make clear which properties of the PC verbs are confined to these verbs specifically, and which properties they share with other verbs that can be used not only lexically but also non-lexically (i.e. functionally or semi-lexically).

The PC verbs are the only verbs of the Afrikaans non-lexical verbal domain that combine with the lexical verb by mediation of *en* (Donaldson 1993, Ponelis 1993, De Vos 2001, Biberauer 2019b).[1] In this sense, the four PC verbs that bring about, among other things, a progressive/durative interpretation of the lexical verb, namely $loop_{progressive}$, *sit*, *staan* and *lê*, form a very clear subgroup of non-lexical verbs in terms of their their morphosyntactic behaviour, as we have seen in the previous chapter that these four verbs almost always require *en* to be present. The fifth PC verb, $loop_{andative}$, only partly belongs to this subgroup of *en*-requiring PC verbs, given that *en* is also very often absent in PC constructions with this verb.

All five PC verbs share the property of (optionally) allowing quirky V2 configurations with a larger subgroup of non-lexical verbs.[2] Biberauer (2019b) gives a list of non-lexical verbs that can occur in quirky V2 configurations for all Afrikaans speakers. It is repeated here in (1).[3]

(1) Afrikaans verbs permitting quirky V2:
- all PC verbs
- motion light verbs: *gaan* 'go', *kom* 'come'
- causative: *laat* 'let'
- benefactive: *help* 'help', *leer* 'teach/learn'

Some examples are given below, all taken from Biberauer (2019b:6-7). As can be seen in the examples, all these non-lexical verbs can occur in V2 position together with the lexical verb, i.e. in a quirky V2 configuration.

(2) Gister ***gaan*** **koop** hy toe vir hom 'n ordentlike motor.
yesterday go buy he PRT for him a proper car
'Yesterday he finally went to go and buy himself a proper car.'

(3) Ons ***help*** **dra** almal die bagasie na die kar.
we help carry all the baggage to the car
'We all help to carry the baggage to the car.'

(4) Die leerders ***leer*** **skryf** hierdie week haikus.
the learners learn write this week haikus
'The learners are learning to write haikus this week.'

[1] They are also referred to as *Indirect Linking Verbs*, based on Ponelis (1968)'s label of *skakelwerkwoorde* 'linking verbs'. *Indirect* refers to the fact that *en* intervenes between the PC verb and the lexical verb. Other Afrikaans semi-lexically used verbs which are directly combined with the lexical verb (like *laat* 'let', *bly* 'remain', *gaan* 'go', or *kom* 'come'), are also referred to as *Direct Linking Verbs* (De Vos 2005:116,118).

[2] See Augustinus and Dirix (2019) for a recent corpus investigation in the degree of optionality of normal/quirky V2 of three non-lexical verbs, namely *laat* 'let', *gaan* 'go' and *kom* 'come'.

[3] Biberauer (2019b:7) notes that for a subgroup of 'innovative' speakers, quirky V2 configurations are also allowed with an additional set of non-lexical verbs, namely *begin* 'begin', *bly* 'stay/remain/keep on', and *probeer* 'try'.

The non-lexical verbs that do not allow quirky V2 configurations, but only normal V2, are the group of modals and auxiliaries (Ponelis 1993, Robbers 1997, De Vos 2005). An illustration of the ungrammaticality of quirky V2 with a modal is given in (5), and for auxiliary *het* in (6).

(5) a. Sy ***moet*** die kinders **help**.
she must the children help
'She must help the children.'
b. *Sy ***moet*** **help** die kinders.
she must help the children
(Robbers 1997:174)

(6) a. Sy ***het*** vandag die boek **gelees**.
she has today the book read
'She read the book today.'
b. *Sy ***het*** **gelees** vandag die boek.
she has read today the book
(Ponelis 1993:326)

The group of PC verbs is thus a distinct subgroup of the non-lexical verbs in that they require *en* to structurally intervene between the PC verb and the lexical verb. $Loop_{andative}$ only partly belongs to this subgroup, as is also allows *en* to be absent. Based on the possibility of quirky V2 configurations, all PC verbs form part of a larger subgroup of non-lexical verbs. Crucially, the most functional-like verbs, i.e. modals and auxiliaries, are not part of this larger subgroup. It thus seems to be the case that all non-lexical verbs that are not (yet) functional, but semi-lexical, allow quirky V2. Table 10.1 provides an overview of the properties of the Afrikaans functional and semi-lexically used verbs that were just discussed. With this overview in place, let us now investigate in which stage of semi-lexicality the various PC verbs can be placed. This is the topic of the next section.

10.3 Stages of semi-lexicality of the PC verbs

In this section, I establish the (degree of) semi-lexicality of the various PC verbs. In the previous case study, we have seen three tests for the semi-lexicality of a verb, namely (i) that it should occur in IPP form when embedded under a perfective auxiliary, (ii) that it does not allow extraposition, and (iii) that it imposes restrictions on the animacy of the subject (in particular in rejecting weather-*it* subjects). Unfortunately, the IPP test does not work for Afrikaans, as the verbal marker *ge-* is generally optional with semi-lexically used verbs. In fact, in section 10.5, I will argue, following Conradie (2012), that the IPP effect does not exist at all in Afrikaans. The details of the will have to wait until that section, but for our current purposes, the consequence is that I cannot use the IPP test to show that the PC verbs are semi-lexical. Let us therefore look

Verbs	*En* present in construction	Quirky V2 allowed
$Loop_{progressive}$	YES	YES
Sit	YES	YES
Staan	YES	YES
Lê	YES	YES
$Loop_{andative}$	OPTIONAL	YES
Gaan	NO	YES
Kom	NO	YES
Laat	NO	YES
Help	NO	YES
Leer	NO	YES
Modals	NO	NO
Auxiliaries	NO	NO

Table 10.1: Properties of Afrikaans functional and semi-lexically used verb

at the two other tests. First, in the examples (7)–(10) it is shown that for all progressive PC verbs, weather-*it* subjects are ungrammatical. Weather-*it* with $loop_{andative}$ is not completely ungrammatical, but not very good either.[4]

(7) ****Dit*** **sit** vandag vir ure en reën.
it sits today for hours and rain
INTENDED: 'It was raining for hours today.'

(8) ****Dit*** **staan** vandag vir ure en reën.
it stands today for hours and rain
INTENDED: 'It was raining for hours today.'

(9) ****Dit*** **lê** vandag vir ure en reën.
it lies today for hours and rain
INTENDED: 'It was raining for hours today.'

(10) ****Dit*** **loop** vandag vir ure en reën.
it runs today for hours and rain
INTENDED: 'It was raining for hours today.'

[4] Theresa Biberauer comments on her rating the sentence with weather-*it* and $loop_{andative}$ as '??' as follows. 'This andative use would be compatible with weather verbs in slightly different circumstances, e.g.:

(i) Toe **loop (en) reën** dit nog boonop!
then walk and rains it even on.top
'And then it went and rained on top of everything else!'

It's just not great with a "straight" present tense as it's naturally part of a lively narrative, rather than a what's-happening-right-now commentary.' (Theresa Biberauer, p.c.).

(11) ??***Dit*** **loop** vandag alweer en reën.
it runs today again and rain
INTENDED: 'It went and rained for hours today.'

The progressive use of *loop* is tested in (10), where the adverbial modifier *vir ure* 'for hours' forces the relevant aspectual reading. Likewise, the andative use of *loop* in (11) is brought about by the absence of this durative adverbial modifier. The ungrammatical examples for *sit* and *staan* are taken from De Vos (2005:142). The three other examples are based on those of De Vos (2005), but since he does not give these exact examples, they were rated by Theresa Biberauer (p.c.).[5,6]

Let us now move on to the other test for semi-lexicality, namely whether the PC verbs block extraposition. As can be seen in the examples in (12-b)–(16-b), this is indeed the case.

(12) a. ... dat hy die boek sit en lees.
... that he the book sit and read
'... that he is (sitting and) reading the book.'
b. *... dat hy sit [die boek en lees].
... that he sits the book and read

(13) a. ... dat hy die boek staan en lees.
... that he the book stand and read
'... that he is (standing and) reading the book.'
b. *... dat hy staan [die boek en lees].
... that he stands the book and read

(14) a. ... dat hy die boek lê en lees.
... that he the book lie and read
'... that he is (lying and) reading the book.'
b. *... dat hy lê [die boek en lees].
... that he lies the book and read

(15) a. ... dat hy die boek loop en lees.
... that he the book walk and read
'... that he is (walking and) reading the book.'
b. *... dat hij loop [die boek en lees].
... that he walks the book and read

(16) a. ... dat hy die boek loop en koop].
... that he the book walk and buy
'... that he goes and buys the book.'

[5] De Vos (2005:142–143) notes that one out of three speakers who judged his sentences did not completely reject the example where *staan* is combined with weather-*it*. He furthermore mentions that in Orange River Afrikaans, this use is acceptable.

[6] See Kocks (1951:30) for examples of spoken language of different regions of the country in which the subject of the PC construction is inanimate. This is also mentioned in Roberge (1994:45).

b. *. . . dat hij loop [die boek en koop].
. . . that he walks the book and buy

In (15-b), the progressive use is tested: one can read a book while walking. In (16-b), the andative use is tested: one more readily goes and buys a book than being in the process of buying one, as the actual buying only takes a second.

Based on the weather-*it* and the extraposition tests, we can establish that all PC verbs have a semi-lexical status. However, before moving on to the next section, it is important to show that they are not all in the same stage of semi-lexicality. In particular, if they all had the same semi-lexical status, we would expect that none of them could embed one of the others in a sentence in which two semi-lexically used verbs co-occur. This is not the case: $loop_{andative}$ can embed the other PC verbs, but the posture verbs cannot embed each other, or $loop_{andative}$ (Biberauer 2019b). Examples of $loop_{andative}$ embedding a posture PC verb are given below; they are all taken from (Biberauer 2019b:11).

(17) Die studente het ***loop*** **sit** en notas maak.
the students have walk sit and notes make
'The students went to sit and make notes (after a lecture).'

(18) Die kat het ***loop*** **staan** en kleintjies hê.
the cat have walk stand and little.ones have
'The cat went and had babies!'

(19) Ons wil net ***loop*** **lê** en TV-kyk.
we want just walk lie and TV-watch
'We just want to go and watch TV.'

A quick check in the *Korpusportaal* corpus (VivA 2016), gives us the numbers of hits for each PC verb embedding the other presented in Table 10.2. Note that in all cases in which *loop* embeds a PC verb, this is the andative use of *loop*: a double marking of progressive/durative aspect is generally ruled out.

	embedded verb			
embedding verb	*Loop*	*Sit*	*Staan*	*Lê*
Loop		9	101	2
Sit	0		0	0
Staan	1	0		0
Lê	0	0	0	

Table 10.2: Embedding properties of Afrikaans PC verbs

As can be seen in the table, the andative use of *loop* can embed all posture verbs, with by far the highest number of hits for it embedding *staan*. The fact that there are but few hits for $loop_{andative}$ embedding *sit* and *lê* is probably due to the fact that these two verbs are less semantically bleached than *staan* (as we have also seen in subsection 9.3.6, and as is also mentioned in Breed (2017) and Biberauer (2019b)) and thus less readily combine with a higher andative semi-lexically used verb. In the table we also see that there is a single hit for *staan* embedding *loop*. The specific example is given in (20).

(20) ... dis hoogtyd dat hy sy pond vleis ***staan*** **loop** en verdien.
... it.is high.time that he his pound meat stands walk and earn
'... it is about time he goes and earns his own money (lit. his pound of meat).' (*Korpusportaal, Leipzig subcorpus*, author unknown)

Though at this time I do not know what to make of this example, the data in the table clearly show that $loop_{andative}$ readily embeds other PC verbs, whereas the other PC verbs cannot embed either each other or $loop_{andative}$. In syntactic terms, this implies that $loop_{andative}$ is merged structurally higher the PC verbs, and that all the other PC verbs have the same, very low, merge position, right above the lexical verb. Based on the data involving weather-*it* subjects and extraposition, we have already established that all PC verbs under investigation here are semi-lexical. The fact that $loop_{andative}$ can embed the other PC verbs and that it seems marginally compatible with weather-*it* subjects, furthermore suggests that this verb is at the second stage of semi-lexicality. All other PC verbs are in the first stage.

In the next three sections, I first discuss the syntactic position of the clause-final auxiliary *het* 'have', the perfect participle marker *ge-*, and *en*. The first is part of the embedded PC constructions tested in the questionnaire study, and given that this verb cannot straightforwardly be analysed as a functional restructuring verb, it needs to be discussed in some more detail. The second is optionally part of the embedded PC constructions, and has been shown to differ significantly from the standard West-Germanic perfect participle marker *ge-* (see amongst others Roberge (1994), De Vos (2003), Conradie (2012)). The third is part of both V2 and embedded PC constructions, and its position in the syntactic structure of Afrikaans PC constructions needs to be determined as well.

10.4 The clause-final auxiliary *het*

In this section, I discuss the morphosyntactic status of the Afrikaans perfective auxiliary *het* 'have'. As this auxiliary is part of the embedded PC constructions tested in this case study, it is important to know what its exact morphosyntactic status is. In Afrikaans, in contrast to Dutch, almost no verbs have a morphologically distinct form for the infinitive (Ponelis (1993), see also (Zwart 2018:4)). One of the very few verbs that do is the possessive use of *het* 'have'. The finite form of the possessive verb is *het*, whereas the infinitival form is *hê*. The contrast is illustrated in (21).

(21) a. Hulle **het** geld.
they have.POS.FIN money
'They have money.'
b. Hulle sal geld **hê**.
they will money have.POS.INF
'They will have money.' (Zwart 2018:4)

The auxiliary *het*, however, does not have a morphologically distinct non-finite form. That is, in contexts in which it is hierarchically embedded under a modal for instance, i.e. contexts where we would expect an infinitival form, *hê* is ruled out, and only *het* can be used. This is illustrated in (22).

(22) a. Hy **het** haar gesien.
he has.AUX.FIN her seen
'He saw her.'
b. Hy sou haar gesien **het** / ***hê**.
he would her seen have.AUX.FIN / have.AUX.INF
'He would have seen her.' (Zwart 2018:5)

As can be seen in the example, both as the finite verb in the sentence (22-a) and as a nonfinite one (22-b), the auxiliary has to appear in the morphological form of *het*. For comparison, the same sentences are given in Dutch in (23), where there is a clear morphological distinction between the finite and non-finite use of the Dutch cognate *hebben* 'have'.

(23) a. Hij **heeft** haar gezien.
he has.AUX.FIN her seen
'He saw her.'
b. Hij zou haar **hebben** gezien.
he would her have.AUX.FIN seen
'He would have seen her.'

Returning to Afrikaans, Conradie (2007, 2018) and Zwart (2018) argue that the clause-final auxiliary *het*, i.e. the auxiliary *het* as it used in (22-b), has grammaticalised into an inflectional ending, one that indicates past tense.[7] According to this analysis, clause-final *het* thus has lost its morphosyntactic status of a real auxiliary verb, and is now an inflectional morpheme, indicating past tense. Given that I will adopt this analysis of clause-final *het*, I now briefly present some pieces of evidence in favour of it. For more discussion and details, I refer the reader to Conradie (2018) and Zwart (2018).

A first indication that there is a very tight relationship between the perfect participle and the auxiliary *het* is the rigid word order between the two elements. While in Dutch it is possible to place the auxiliary *hebben* both be-

[7]For the purposes of this case study, only the morphosyntactic status of clause-final *het* matters, and I therefore do not take a position on the status of the V2 auxiliary *het* here, see Conradie (2007, 2018), Zwart (2018) for discussion.

fore and after the perfect participle in clause-final position (24), in Afrikaans auxiliary *het* can only follow the perfect participle (25) (Conradie 2007, 2018, Zwart 2018).

(24) ... dat hij **gewerkt** ***heeft*** / ***heeft*** **gewerkt**.
... that he work.PTCP has.AUX / has.AUX work.PTCP
'... that he has been working.'

(25) ... dat hy **gewerk** ***het*** / ****het*** **gewerk**.
... that he work.PTCP has.AUX / has.AUX work.PTCP
'... that he has been working.'

An indication for *het* being an inflectional morpheme rather than a verb is the fact that it is inseparable from the perfect participle with which it forms the periphrastic perfect. In particular, in verb clusters with a hierarchical order of modal$_1$, passive auxiliary$_2$ and perfect participle$_3$, the perfect participle can be separated from the passive auxiliary (26), but in clusters with the same hierarchical order in which the auxiliary is *het*, the perfect participle cannot be separated from *het* (27).

(26) a. ... dat die huis **moes**$_1$ **gebou**$_3$ ***word***$_2$.
... that the house must.MOD build.PTCP be.PASS.AUX
b. ... dat die huis **gebou**$_3$ **moes**$_1$ ***word***$_2$.
... that the house build.PTCP must.MOD be.PASS.AUX
'... that the house had to be built.'

(27) a. ... dat die huis **moes**$_1$ **gebou**$_3$ ***het***$_2$.
... that the house must.MOD build.PTCP have.AUX
b. *... dat die huis **gebou**$_3$ **moes**$_1$ ***het***$_2$.
... that the house build.PTCP must.MOD have.AUX
'... that they had to build the house.' (Conradie 2018:3)

Another illustration of the tight relationship between the perfect participle and the auxiliary *het* is the fact that in non-finite constructions, these two elements cannot be separated by *te* 'to', whereas in all other cases, the perfect participle and the auxiliary or semi-lexically used verb are always separated by *te*. The contrast is illustrated in (28) and (29).

(28) a. ... om **gebou** **te** ***word*** / ***wees*** / ***kom***.
... for build.PTCP to be.PASS.*aux* / be.AUX / come
'... to become/be/get built.'
b. *... om **te gebou** ***word*** / ***wees*** / ***kom***.
... for to build.PTCP be.PASS.*aux* / be.AUX / come

(29) a. *. . . om **gebou** **te** ***het***.
. . . for build.PTCP to have.AUX
b. . . . om **te gebou** ***het***.
. . . for to build.PTCP have.AUX
'. . . to have built.' (Conradie 2018:3)

A phonological indication for the morphosyntactic status of *het* as an inflectional morpheme is the fact that it is frequently reduced to *'t*, both in writing and in speech (Conradie 2018:5).

Based on all these indications, Conradie (2018) concludes that clause-final *het* has grammaticalised into an inflectional morpheme, more specifically a suffix. He furthermore argues that the function of this suffix is to express past tense. His reasoning goes as follows. First, an enabling factor in the development of *het* was a change that took place early in the 19_{th} century, in which unaccusative verbs no longer took the auxiliary *is/wees* 'be', but selected *het* 'have' instead. As a result, whereas in Dutch unaccusative verbs are embedded under the auxiliary *zijn* 'be', and all others under *hebben* 'have' (30), from the 19_{th} century onwards, all Afrikaans verbs were embedded under *het* (31) (Conradie 2018:7).

(30) a. . . . dat ik ***ben*** / ****heb*** **gegaan**.
. . . that I am.AUX / have.AUX go.PTCP
'. . . that I went.'
b. . . . dat ik ****ben*** / ***heb*** **gekookt**.
. . . that I am.AUX / have.AUX cook.PTCP
'. . . that I cooked.'

(31) a. . . . dat ek **gegaan** ***het*** / ****is***.
. . . that I go.PTCP have.AUX / am.AUX
'. . . that I went.'
b. . . . dat ek **gekook** ***het*** / ****is***.
. . . that I go.PTCP have.AUX / am.AUX
'. . . that I cooked.'

This resulted in a huge increase in the frequency of the auxiliary *het* relative to the auxiliary *is/wees*. A second development that played a crucial role in the grammaticalisation of *het* is the loss of the preterite past, also at the beginning of the 19_{th} century (Conradie 2018:7).[8] This loss caused the periphrastic perfect to be the only means of indicating past tense, also in cases in which the event is not related to the present, i.e. in contexts where previously the preterite form was used (Conradie 2012:141–142). This means that while in Dutch there is a distinction between the past tense and the periphrastic perfect (32), in

[8] Only the modals have kept their preterite form: *sou* 'would', *moes* 'must', *wou* 'would want', and, albeit somewhat archaically, *mog* 'might'. These verb forms are most often used in a modal function, however, and not to indicate past tense (Conradie 2012:141). See Loubser (1960) and Conradie (1999) for a detailed description of this development in the history of Afrikaans.

Afrikaans, the only way of referring to the past is by using the periphrastic perfect (33).

(32) a. ... dat hij **zong**.
... that he sing.PRT.PAST
'... that he sang.'

b. ... dat hij **heeft gezongen**.
... that he has.AUX sing.PTCP
'... that he has sung.'

(33) a. *... dat hy **zong**.
... that he sing.PRT.PAST

b. ... dat hy **gesing het**.
... that he sing.PTCP has.AUX
'... that he sang/has sung.' (Conradie 2018:2)

The fact that the periphrastic perfect was now the only means of refering to a past event further increased the use of the auxiliary *het* in combination with a perfect participle, and this led to the reanalysis of the auxiliary *het* as a past tense marker (Conradie 2018:7). This analysis is also corroborated by frequency data of the past tense of the verb *wees* 'to be'. This is one of the few verbs that can still be embedded under the auxiliary *is/was* 'be' rather than *het*. However, as Conradie (2018) shows by means of a corpus study, in clause-final position, even *wees* is embedded under *het* rather than *is/was*. This shows that in clause-final position, *het* is the inflectional morpheme used to express past tense across the board. Conradie (2018)'s data, extracted from the *Taalkommissiekorpus 1.1*, are given here in Table 10.3.

Auxiliary	V2 position	clause-final position	Total
Was	*was ... gewees*	*gewees was*	
Hits	190 (95,0%)	2 (5%)	192 (100%)
Is	*is ... gewees*	*gewees is*	
Hits	8 (100%)	0 (0%)	8 (100%)
Het	*het ... gewees*	*gewees het*	
Hits	1 (< 0,1%)	3760 (> 99,9%)	3760 (100%)

Table 10.3: Corpus results (Conradie 2018:3) on embedding of *gewees* 'been'

As can be seen in the table, by far the most frequent auxiliary of *gewees* in clause-final position is *het*. In root clauses, the auxiliary in V2 position that embeds clause-final *gewees* is most frequently *was* 'were'. These data thus show that in clause-final position, the only means of expressing past tense is *het*, regardless of the auxiliary the participle is usually embedded under in root clauses.

Summing up, we have seen that the Afrikaans auxiliary *het* is no longer a real auxiliary verb in clause final position, but rather an inflectional morpheme. More specifically, it is a suffix that attaches to the participle, and that marks past tense. With the morphosyntactic status of clause-final *het* established, we now move on to the morphosyntactic status of another element that can occur in PC constructions, namely the perfect participle marker *ge-*.

10.5 The perfect participle marker *ge-*

In this section, I discuss the Afrikaans perfect participle marker *ge-*. First, I show that Afrikaans *ge-* is not identical to the European West-Germanic (henceforth 'EWG') perfect participle marker *ge-*. Second, I argue that the optionality of *ge-* in embedded PC constructions should not be seen as an optional IPP effect. Third, I propose to analyse the optionality of *ge-* in embedded PC constructions as caused by two factors. The first factor is that *ge-* can attach to phrasal material. The second factor is that the stress pattern of the combination of the PC verb and lexical verb is the type of stress pattern which renders *ge-* optional in Afrikaans in general (Conradie 2012). The fact that both *ge-* and *het* indicate past tense, might be an additional factor making *ge-* optional in PC constructions containing *het*.

10.5.1 Afrikaans *ge-* is not identical to European West-Germanic *ge-*

The perfect participle marker *ge-* is different from *ge-* in EWG in several ways. A first difference is that in EWG languages that have the perfect participle marker *ge-*, this marker is part of a circumfix, namely *ge-*...-t/*d* or *ge-*...-*en* (Zwart 2007:77), whereas in Afrikaans only *ge-* marks a verb as a perfect participle (Conradie 2012:131).[9] Two examples of the circumfix from Standard Dutch are given in (34) and (35).

(34) Ik heb het ***ge***-zeg-***d***.
I have it GE-say-D
'I have said it.'

(35) Ik heb ***ge***-lop-***en***.
I have GE-walk-EN
'I have walked.'

The Afrikaans counterparts of the perfect participles in these sentences are given in (36) and (37).

[9]See Conradie (1979) for diachronic work on the loss of the circumfix in Afrikaans.

(36) Ek het dit ***ge*-sê**.
I have it GE-say
'I have said it.'

(37) Ek het ***ge*-loop**.
I have GE-walk
'I have walked.'

As can be seen in the examples, the Afrikaans perfect participles consist of the perfect participle marker *ge-* plus the verbal stem. *Ge-* thus no longer forms part of a circumfix, and is an autonomous morpheme that by itself creates perfect participles (Conradie 2012:131).[10]

A second difference between EWG *ge-* and Afrikaans *ge-* is the fact that the latter can be used to express past tense in periphrastic constructions, whereas the former cannot. That is, the Afrikaans periphrastic perfect has broadened into a past tense.[11] The Afrikaans periphrastic perfect construction can therefore both refer to telic and atelic events, as shown in (38).

(38) Ek **het** die boek **gelees**.
I have the book read.PTCP
'I have read the book (telic/completed).'
'I read the book (atelic/uncompleted).'

In contrast, the periphrastic perfect in Dutch most naturally refers to a completed event (39-a), while the simple past can be used to indicate the the event is not yet completed (39-b).

(39) a. Ik **heb** het boek **gelezen**.
I have the book read.PTCP
'I have read the book (telic/completed).'
b. Ik **las** het boek.
I read.PAST the book
'I read the book (atelic/uncompleted).

(Conradie 2012:131) takes this to mean that Afrikaans *ge-* has undergone resemanticisation, and is now semantically related to a broad notion of past, rather than just to perfect.

A third difference between the two *ge-*'s is that in Afrikaans *ge-* is the only verbal marker left in the verbal system, i.e. there is a single morphological split between the base form of a verb (e.g. *ry* 'ride') and its participial form (e.g. *gery*

[10]Note that Ablaut also no longer plays a role in the formation of Afrikaans perfect participles. Whereas in Dutch and German, strong verbs undergo Ablaut in the formation of the past participle in combination with the perfect participle marker *ge-*, the Afrikaans cognates follow the general Afrikaans pattern of *ge-* plus verbal stem (e.g. Dutch *rijden–gereden* 'ride–ridden' versus Afrikaans *ry–gery*, Dutch *brengen–gebracht* 'bring–brought' versus Afrikaans *bring–gebring* (Conradie 2012:131).

[11]See De Vos (2003:524) and Conradie (2012:135). See Roberge (1994:68) on the fact that *ge-* was already a past tense marker in Cape Dutch creole.

'ridden'). In contrast, in EWG, there are other verbal markers as well, such as the infinitival suffix *-en* or the finite verbal endings. This leads to a three-way verbal distinction between finite (e.g. *rijd(t)* 'ride'), infinitival (e.g. *rijden* 'to ride') and participial forms (e.g. *gereden* 'ridden') (Conradie 2012:135). The fact that *ge-* is the only morphological means of marking a contrast between verb forms makes it a likely candidate for acquiring additional functions other than its original one.

A fourth difference is the complementary distribution with other prefixes. In Dutch, *ge-* is in complementary distribution with verbal prefixes such as *be-*, *ver-*, *ont-*, *vol-* and *her-*. For example, the perfect participle of *be-ginnen* 'start' is *be-gonnen* rather than *ge-be-gonnen*, and similarly, the perfect participle of *ont-moeten* 'meet' is *ont-moet* rather than *ge-ont-moet.* In Afrikaans, however, this complementary distribution, while mostly still present as a normative rule of grammar, is often ignored in the spoken language (Conradie 2012:140). Some examples from a corpus of spoken Afrikaans (Kroes 1982) (given in (Conradie 2012:140)) are *ge-be-waar* 'preserved', *ge-ont-moet* 'met', *ge-ont-werp* 'designed' and *ge-vol-tooi.* It is important to note here that *ge-* always appears linearly to the left of the other verbal prefixes (i.e. *ge-ont-moet* but not **ont-ge-moet*). This is an indication that *ge-* is in a hierarchically higher position and and hence further removed from the verbal stem than the other verbal prefixes.

In summary, we have seen four pieces of evidence in favor of the assumption that Afrikaans *ge-* is not identical to EWG *ge-*. Two of them suggest that Afrikaans *ge-* is morphologically more autonomous than EWG *ge*, namely the fact that it is not part of a circumfix, and that it is situated further removed from the verbal stem than other verbal prefixes. The other two pieces of evidence suggest that Afrikaans *ge-* has a more unique status in the verbal system than its EWG cognate: (1) it is the only verbal marker left in Afrikaans, (2) it can be used to express past tense in addition to being a perfect participle marker. Having established that Afrikaans *ge-* is clearly different from EWG *ge-*, the question arises whether Afrikaans is subject to the IPP-effect, an EWG phenomenon whereby the *ge*-marker plays a central role. This question is the topic of the next subsection, where I propose that Afrikaans IPP seems restricted to a limited set of verbs compared to EWG. IPP does not seem to apply to the Afrikaans restructuring verbs which are innovative verbs in the language, i.e. verbs that were not directly inherited from Middle Dutch. The Afrikaans PC verbs fall in the class of innovative verbs, which suggests that the optionality of *ge-* in embedded PC constructions is not caused by an optionality in the IPP effect.

10.5.2 Innovative Afrikaans verbs do not have IPP

I have already introduced and illustrated the IPP effect in chapter 2. It is the phenomenon whereby a bare or infinitival form of a verb surfaces in contexts where selectional requirements dictate that a perfect participle should. An example from Dutch is given again in (40).

(40) Ik **ben**$_1$ lang ***blijven***$_2$ / ****gebleven***$_2$ **lezen**$_3$.
I am.AUX long remain.IPP / remain.PTCP read
'I've continued reading for a long time.'

A perfective auxiliary, such as V1 *ben* 'am' in (40), normally selects a perfect participle. However, V2 *blijven* 'remain' in (40) obligatorily appears in the IPP form rather than as a perfect participle. Recall that this only happens when this verb itself selects an infinitive of its own (V3 *lezen* 'read' in (40)), i.e. in two-verb clusters in which V1 is a perfective auxiliary V2 always surfaces as a perfect participle (41).

(41) Ik **ben**$_1$ er lang ****blijven***$_2$ / ***gebleven***$_2$.
I am.AUX there long remain.IPP / remain.PTCP
'I've stayed there for a long time.'

The IPP effect exists in Dutch and its dialects and in German and its dialects (Schmid 2005), but not in the other West-Germanic languages, namely English, Frisian and Yiddish (Hinterhölzl 2009). In much of the literature on the IPP effect, or on Afrikaans syntax in general, it is assumed that Afrikaans also shows the IPP effect, albeit optionally (see among others Du Plessis Scholtz (1963), Ponelis (1979), Donaldson (1993), Robbers (1997), De Vos (2001), Schmid (2005), Zwart (2007), Augustinus and Dirix (2013), Dirix et al. (2020)). Donaldson (1993) notes, though, that the presence of *ge-* in typical IPP contexts is strongly preferred in colloquial Afrikaans. For example, in a sentence like (42), in which the IPP effect is supposed to be optional in Afrikaans, the non-IPP configuration (i.e. with *ge-* on *bly* 'remain') is strongly preferred over the IPP configuration (without *ge-* on *bly*) in colloquial Afrikaans according to Donaldson (1993:225–226).

(42) Ek **het**$_1$ lank ***bly***$_2$ / ***gebly***$_3$ **lees**$_3$.
I have long remain.IPP / remain.PTCP read
'I've continued reading for a long time.'

However, not all scholars working on Afrikaans agree that IPP, even if only optional, is still an active part of Afrikaans grammar. For example, Ponelis (1993:413) claims that Afrikaans only has "a residue" of the EWG IPP effect. Taking it one step further, De Schutter (2001:205) argues that the fewer and often optional attestations of IPP in Afrikaans cannot be seen as a residue of the EWG IPP effect, but that it shows that the phenomenon has developed in a different direction in this language. Conradie (2012:142) claims that the IPP effect is simply no longer a part of Afrikaans grammar. I assume that IPP is still part of Afrikaans grammar, but works on a more restricted class of restructuring verbs compared to Dutch. The PC verbs do not seem to fall in the restricted class of IPP verbs in Afrikaans.

Evidence for this assumption comes from the fact that the there is a clear difference in the frequency of the IPP form (i.e. the bare form without *ge*-) and the perfect participle form (i.e. the bare form plus *ge*-) among different subclasses of Afrikaans verbs that can occur as V2 in three-verb clusters in which V1 is a perfective auxiliary (i.e. in the prototypical IPP contexts in EWG). This is shown in a recent corpus study by Dirix et al. (2020), in which they investigate the optionality of Afrikaans IPP with several subclasses of verbs. They use the *Taalkommissie* corpus (54 million words, written Afrikaans) and the Afrikaans section of Wikipedia (see Dirix et al. (2020) for the details of their methodology). The results of their corpus search are summarised in Table 10.4.[12,13]

As can be seen in the table, there are a number of verbs for which the total number of hits is so low that we do not want to base any conclusions on them. This holds for specifically for *weet* 'manage', *blyk* 'turn out', *maak* 'make', *hoor* 'hear', *leer* 'teach' and *loop* 'walk' (all less than 13 hits), and to a lesser extent for *ophou* 'stop', *leer* 'learn', and *lê* (all less than 40 hits). Given that a subgroup of these verbs has a low (and sometimes very low) percentage of IPP cases, this is a pity, because we now do not know whether this low incidence of IPP is significant or not.

With this caveat in mind, the table shows us that for most verbs, IPP is by far the most common option. Verbs that have a slightly lower total percentage of IPP compared to others are *aanhou* 'continue' and *ophou* 'stop', and the PC verb *lê*. The two verbs which a much lower total percentage of IPP are the other two posture PC verbs, *sit* and *staan*. Note that all these verbs are innovations of Afrikaans. That is, *aanhou* and *ophou* are not clustering verbs in Dutch, and, as I discuss in the next section, the Afrikaans PC verbs are not a direct borrowing from Middle Dutch either. This suggests that IPP in Afrikaans is an active rule for the IPP verbs that were inherited from Dutch (cf. Ponelis (1993)), but that this rule was not extended to the class of innovative clustering verbs.

In order to further explore the frequency distribution of IPP versus perfect participle for a number of the verbs in Table 10.4, Liesbeth Augustinus and I conducted a corpus search in a bigger Afrikaans corpus, namely *Korpusportaal*

[12] I have merged their Table 3 with their Table 4. Furthermore, I have combined the numbers of the *Taalkommissie* corpus and *Wikipedia*, which in their tables are presented separately, as I am not concerned here with the type of corpus as a possible effect on the occurrence of IPP. I have also left out the column 'IPP + *te*', but I have indicated in the IPP-column if a verb had hits involving a *te*-complement. This is the case for all hits with *weet* 'know', and a number of hits with *blyk* 'turn out'.

[13] I do not discuss the class of modals here. These verbs can occur in IPP form or preterite form (i.e. *moes* for *moet*, *kon* for *kan*, et cetera) in embedded verb clusters. These clusters can also occur in different word orders (e.g. *het kon/kan werk* and *kon/kan gewerk het*). The interested reader is referred to Dirix et al. (2020) for discussion and corpus analysis of the different possible verb cluster configurations and morphological forms modals can occur in.

Type	Verb	IPP form	perfect participle	Total	Total %IPP
Aspectual	*begin* 'begin'	2176	2	2178	99,91%
	gaan 'go'	1053	0	1053	100,00%
	kom 'come'	792	17	809	97,90%
	bly 'remain'	357	1	358	99,72%
	aanhou 'continue'	59	9	78	88,46%
	ophou 'stop'	26	6	32	81,25%
Subj. control	*probeer* 'try'	747	1	748	99,87%
	durf 'dare'	39	1	40	97,50%
	leer 'learn'	27	11	38	71,05%
	weet 'manage'	3 (+*te*)	0	3	100,00%
Evidential	*blyk* 'turn out'	6 (+*te*)	6	12	50,00%
Causative	*laat* 'let'	2009	2	2011	99,90%
	maak 'make'	2	5	7	28,57%
Perception	*sien* 'see'	139	8	147	94,56%
	hoor 'hear'	4	0	0	100,00%
Benefactive	*help* 'help'	181	11	192	94,27%
	leer 'teach'	2	1	3	66,67%
PC verbs	*loop* 'walk'	1	5	6	16,67%
	sit 'sit'	50	60	110	45,45%
	staan 'stand'	48	49	97	49,48%
	lê 'lie'	30	5	35	85,71%

Table 10.4: Corpus results (Dirix et al. 2020) on Afrikaans IPP

(VivA 2016).[14,15] The verbs that we investigated are *maak* 'make', *hoor* 'hear', *help* 'help', *bly* 'remain', *ophou* 'stop', and the four PC verbs.[16] The results are presented in Table 10.5.

As can be seen from the table, there are more hits for all verbs than there were in Table 10.4, which is a welcome result. The only verb that still has very low frequencies is causative *maak* 'make'. The data show that for the three verbs that are also IPP verbs in Dutch, namely *bly* 'remain', *hoor* 'hear' and *help* 'help', the total percentages of IPP are very high (all above 90%), as was the case in Table 10.4. The Afrikaans innovative clustering verbs, *ophou* 'stop', *maak* 'make' and the PC verbs, all have lower total frequencies of IPP (all below 77%). Note that especially the posture PC verbs have a total IPP frequency below 50%. These findings thus support the hypothesis I presented

[14]The results of this corpus search are already partly presented in chapter 8. The methodology we used is identical to the one presented in that chapter, as is the make-up of the corpus.

[15]I am grateful to Liesbeth Augustinus for allowing me to present our data here.

[16]Recall from section 8.3 that all hits for the PC verb *loop* 'walk' were annotated by two native Afrikaans speakers for progressive vs. andative aspect. However, as there were quite a few unclear cases no split is made between the two uses of *loop* in Table 10.5.

Type	Verb	IPP form	perfect participle	Total	Total %IPP
Aspectual	*bly* 'remain'	2134	6	2140	99,71%
	ophou 'stop'	498	310	808	61,63%
Causative	*maak* 'make'	10	4	14	71,43%
Perception	*hoor* 'hear'	545	26	581	93,89%
Benefactive	*help* 'help'	539	50	589	91,51%
PC verbs	*loop* 'walk'	83	28	109	76,14%
	sit 'sit'	220	235	455	48,35%
	staan 'stand'	155	191	346	44,79%
	lê 'lie'	113	136	249	45,38%

Table 10.5: Corpus results from *Korpusportaal* on Afrikaans IPP

above, the Afrikaans IPP effect is restricted to a smaller class of verbs compared to Dutch, to which the Afrikaans innovative restructuring verbs do not belong. However, the fact that *ge-* optional in embedded constructions with the innovative Afrikaans verbs still needs to be accounted for. I present my take on this in the next subsection.

10.5.3 *Ge-* in the current analysis

In this subsection, I propose that *ge-* in embedded PC constructions is the result of the combination of two factors. The first is that *ge-* is a clitic that can attach to phrasal material. The second factor is the existence of a phonological rule which makes *ge-* optional when a verb or combination of two verbs has an unstressed-stressed pattern. I discuss each factor in turn.

I follow Conradie (2012) in assuming that Afrikaans *ge-* is a clitic that can attach to phrasal material. He argues that Afrikaans semi-lexically used verbs can form a phrasal participle with the lexical verb. According to him, the scope of *ge-* can extend beyond word boundaries, and as such it is similar to the English possessive *'s* as in *the woman with the long hair's suitcase*. He bases this claim on examples such as the one in (43), in which the V4+V5 complex can be scrambled as a unit, while scrambling in Afrikaans verb clusters is normally only allowed with participles.

(43) a. Ons hoop dat die rokery deur hulle **sal**$_1$ **moet**$_2$
we hope that the smoke.NOM by them shall must
(ge)-laat$_4$ ***staan***$_5$ **word**$_3$.
GE-let stand become

b. Ons hoop dat die rokery deur hulle **sal**$_1$ ***(ge)-laat***$_4$
we hope that the smoke.NOM by them shall GE-let

staan$_5$ **moet**$_2$ **word**$_3$.
stand must become

c. Ons hoop dat die rokery deur hulle ***(ge)-laat***$_4$ ***staan***$_5$
we hope that the smoke.NOM by them GE-let stand
sal$_1$ **moet**$_2$ **word**$_3$.
shall must become
'We hope that they will have to stop smoking (lit. that they will have to let the smoking be stood).' (Conradie 2012:142)

In all three examples, we can see that V4 and V5 *(ge)laat staan* stay together while being scrambled in different positions. Based on this, Conradie (2012) takes V4 and V5 to be a complex participle, in this case a passive participle selected by V3 *word*. The example furthermore shows that *ge-* can optionally appear on V4 *laat*. Assuming that V4 indeed forms a complex participle with V5, this is an indication that *ge-* can also attach to phrasal material (a phrasal participle in this case).

The optionality of *ge-* in Afrikaans PC verbs seems to follow from a phonological rule, combined with the fact that *ge-* can attach to phrasal material. Conradie (2012:142–143) shows that there is a phonological condition on the presence or absence of *ge-* on Afrikaans participles in general. In monosyllabic verbs, *ge-* has to be present on the participle. For example, the participle of *laat* 'let' is *gelaat*. In polysyllabic words, *ge-* is obligatory on the participle when the main stress lies on the first syllable. Examples are *ántwoord-geántwoord* 'answer-answered', *hárdloop-gehárdloop* 'jog-jogged', and *rááđpleeg-geráádpleeg* 'consult-consulted'. *Ge-* is optional in participles when the main stress does not lie on the first syllable, for example *probéér-(ge)probéér* 'try-tried'. Crucially, this phonological condition of *ge-* also holds for the combination of the semi-lexical and the lexical verb in verb clusters. The lexical verb always bears the main stress of the two verbs, which results in an unstressed-stressed pattern. This pattern makes *ge-* optional, just as it is within a single word such as *(ge)probéér*. Examples include complex participles such as *(ge)laat léés* 'let read' and *(ge)bly stáán* 'remain standing'. Note that this extends to the PC verbs as well, in which the main stress lies on the lexical verb: *(ge)sit en léés* 'sit and read'. In other words, the stress patterns of embedded PC constructions renders *ge-* optional.

In the previous subsections we have seen that *ge-* in Afrikaans is closely connected to past tense. I therefore take *ge-* to be a past tense marker. Recall from section 10.4 that the clause-final auxiliary *het* also has the status of a past tense marker. This means that these two elements carry the same or at least a very similar meaning. The presence vs. absence of *ge-* in Afrikaans three-verb clusters can thus be seen as two variants of the same construction. I therefore hypothesise that the presence of *het* in embedded PC constructions might be an additional factor making *ge-* optional in these structures.

10.6 *En* in PC constructions

In this section, I discuss the morphosyntactic status of *en* in Afrikaans PC constructions. I also discuss different views on the diachronic development of Afrikaans PC constructions, in order to sketch a picture of the complexity underlying the question of the morphosyntactic status of *en*. In the following four subsections, I discuss the proposals by Kocks (1951), Roberge (1994), De Vos (2005) and Biberauer (2019b) respectively. In subsection 10.6.5, I present my own account.

10.6.1 Kocks (1951)

According to Kocks (1951:12–13), *en* developed from a coordinator into a subordinator. This development was caused by the very frequent combination of a PC verb and a lexical verb in a coordination construction. This resulted in a reanalysis, whereby the relation between the PC verb and the lexical verb was reinterpreted as asymmetric rather than symmetric. As a result, *en* was also reanalysed as a subordinator. This syntactic reanalysis was followed by the phonetic weakening of *en*. That is, it became an unstressed element, and often forms a phonological unit with the PC verb. Furthermore is often pronounced as /ə/, rather than /ən/. For example, the Afrikaans speakers of Calitzdorp (a location in the Eastern Cape province) remark that PC constructions in the written language are of the type given in (44), in which *en* is a separate element, whereas in the spoken language these speakers use constructions like the one in (45), in which *en* has been reduced to a schwa that is attached to the PC verb.

(44) staan **en** praat
stand and talk
'be (standing and) talking'

(45) stan-**e** praat
stand-e talk
'be (standing and) talking' (Kocks 1951:12)

Based on these findings, Kocks (1951) argues that "nowadays" (i.e. in 1951) *en* is a sort of epenthetic vowel (which she calls *oorgangsklank* lit. 'transition sound'), the semantics of which is almost completely empty. She furthermore speculates that it remained part of the PC construction due to "considerations of rhythm", i.e. it enhances the prosodic contour of the PC construction. This remains a stipulation in her account, however, and it is not very likely in light of the diachronic development of PC constructions, where the version without precedes the one with *en*. Note that from a phonological point of view Kock's account is not very plausible either, for several reasons. First, all PC verbs can be combined with *en*, including a verb like *lê*, which does not end in a consonant. Given that epenthetic vowels are typically inserted between two consonants, it is unclear how such a vowel could improve the prosodic contour of a *lê*-PC-construction. Second, many other semi-lexically used verbs in Afrikaans, like *kom* 'come', *gaan* 'go', and *laat* 'let', also end in a consonant, and would thus be possible candidates for the insertion of an epenthetic vowel when combined with a lexical verb. However, with these other semi-lexically used verbs, such a vowel is never present. Third, even though there is no phonological literature that I am aware of about epenthesis in Afrikaans, in Dutch epenthesis is specifically used after sonorants (for example, /mɛlək/ for *melk* 'milk'). Only one of the four Afrikaans PC verbs (i.e. *staan*) end in a sonorant, however.[17]

In fact, Kocks (1951) herself also notes that even though in 1951, PC constructions with *en* are the standard written form, in earlier written sources there was much uncertainty about the 'correct' use (i.e. with or without *en*). She did a diachronic corpus study, and found that *en* was only consistently used in PC constructions from 1929 onwards (from D.F. Malberbe's *Hans-die-Skipper*). She also refers to Du Toit (1905:18), who only mentions PC constructions without *en* in his dissertation on Afrikaans syntax. It thus seems that somewhere between the start of the 20^{th} century and the middle of this century, there must have been a rapid increase in PC constructions with *en*. Furthermore, there must also have been a normative push towards PC constructions with *en* during that period. This is confirmed by Kocks (1951), who mentions that only PC constructions with *en* are correct (*beskaaf* 'civilised', in her own words) at schools. In other words, PC constructions with *en* were only consistently used in Afrikaans prose from the late 1920s onwards. Furthermore, the use of *en* in PC constructions was actively promoted at least in the 1950s. All of this suggests that *en*-less PC constructions existed in Afrikaans prior to the constructions with *en*, and that the latter is possibly a product of standardisation and normative pressure. In line with this is the observation of De Klerk (1968:232) that in 1968, the deletion of *en* in Afrikaans PC constructions was in decline throughout the entire Afrikaans-speaking population. This too suggests that before that period, PC constructions without *en* were well attested in the entire country.

A final note on Kocks (1951)'s discussion of Afrikaans PC constructions: she briefly mentions an observation from Rademeyer (1938), who says that PC constructions with *en* are not used by 'Coloured' Afrikaans speakers, and that they only use the construction without *en*.[18] She does not elaborate more on the role of the *en*-less version in the development of Afrikaans PC constructions in general. However, this topic is discussed in more detail in Roberge (1994), to which I now turn.

[17] Many thanks to Edoardo Cavirani for discussing the phonological implications of Kocks (1951)'s account with me.

[18] I am aware of the highly problematic nature of this label, and use it here only to refer to Rademeyer (1938)'s own phrasing.

10.6.2 Roberge (1994)

Roberge (1994) discusses three views on the origins of Afrikaans PC constructions, the 'Netherlandicist' view, the 'creolism' view, and his own view, in which the current Afrikaans PC constructions are the result of a 'compromise' between the Dutch construction and the verb serialisation constructions of the basilect. I discuss all these three views below.[19]

According to the 'Netherlandicist' view, Afrikaans PC constructions have the verbal hendiadys construction of Early Modern Dutch as their direct ancestor. In Middle Dutch, posture verbs were combined with a lexical verb in a hendiadys construction (see amongst others Stoett (1923), Van Pottelberge (2002), Haslinger and Koppen (2002-2003)). An example from Middle Dutch is given in (46), and one from Early Modern Dutch in (47).[20]

(46) Hij **stond ende dachte**.
he stood and thought
'He stood and thought.' (Stoett (1923), paragraph 13)

(47) ... dat jij hier **lecht en tabackt**.
... that you here lie and smoke.tobacco
'... that you lie here and smoke tobacco.' (Bredero, *Griane*, page 105)

Advocates of the 'Netherlandicist' view (Bouman (1926:25), Kempen (1965:85-86) and Raidt (1983:180-181)) argue that the verbal hendiadys of Middle and Early Modern Dutch was exported to the Cape of Good Hope, while the younger Early Modern Dutch construction with *te*, which made its entrance in the Dutch language in the 17th century, was not. Thus, the Afrikaans PC construction was directly based on the Middle and Early Modern Dutch verbal hendiadys construction. However, there are several problems with this hypothesis. For example, as we have also seen above in Kocks (1951)'s corpus study, *en*-less PC constructions were very frequent before 1900. If the Afrikaans PC construction directly originates from the Early Modern Dutch verbal hendiadys, this is quite unexpected. Furthermore, as Roberge (1994:54) notes, hardly any occurrences of verbal hendiadys can be found in the existing sources from Cape Dutch from the 17th, 18th and 19th century. In other words, in the early centuries of the Afrikaans language, there was no Dutch verbal hendiadys input to borrow from and implement into the language. For more arguments against the 'Netherlandicist' view on the origins of Afrikaans PC constructions, I refer to Roberge (1994:54-58), but let me finish the argumentation against this view by a quote from his paper:

> Particularly awkward for the Netherlandicist position is the curious relative chronology for what was supposedly an ordinary feature

[19] This subsection is a summary of Roberge (1994). I would like to emphasise here that all what is mentioned is thanks to his diachronic research, not my own.

[20] Both examples are taken from Roberge (1994:50).

> in the Dutch 'volkstaal' [language of the people, red.] of the seventeenth century. Bouman (1926:7) considered the Afrikaans hendiadys "[een] van de voornaamste afwijkingen van Nederlands op syntakties gebied." [*one of the most remarkable divergences from Dutch from a syntactic point of view* (translation added by me)]. Yet, unlike other salient Africanderisms, the hendiadys construction as we know it in Modern Afrikaans seems to have passed beneath the notice of the early lexicographers, grammarians, and dillettante observers. This fact alone suggests that the Afrikaans verbal hendiadys must be somewhat younger than the Nederlandicist position would have us believe, and that its antecedents may lie elsewhere, namely, in basilectal forms of Afrikaans. (Roberge 1994:57)

In other words, given that the Afrikaans PC construction is one of the most clear divergences from Dutch syntax (according to Bouman (1926)), it is very suspicious that it has not received any attention in the literature on Afrikaans in earlier periods. I agree with Roberge (1994) that this is a clear indication that the Afrikaans PC construction was simply not yet used in those early periods, or at least not by the speakers that would have been on the radar of lexicographers and grammarians.

Another view on the origins of Afrikaans PC construction is that it is a 'creolism' (Roberge 1994:58), i.e. that its origin lies in the basilect rather than in Dutch. A piece of evidence in favour of this view is the observation of Bosman (1923:83) that PC constructions were often used in 'creole' ('Coloured') Afrikaans (henceforth *basilectal Afrikaans*, following Roberge (1994)), but much less so in what he called 'beskaafde Afrikaans' (lit. 'civilized Afrikaans', i.e. spoken by white Afrikaners).[21] The peculiar quirky V2 property of Afrikaans PC constructions is also thought to have come from basilectal Afrikaans. For example, observations of quirky V2 with PC verbs in basilectal Afrikaans have been made in Du Toit (1905:95). Two such examples are given in (48)–(49).

(48) Hij **stan melk** die koeie elke dag.
he stands milk the cows every day
'He is milking the cows every day.' (Du Toit 1905:95)

(49) Jij **stan speul** als jij moet werk.
you stand play when you must work
'You're playing while you should be working.' (Du Toit 1905:95)

An argument for the idea that Afrikaans PC constructions are a 'creolism' that has been put forward by many authors is that in basilectal Afrikaans, PC verbs look more like verbal prefixes rather than actual verbs, which in turn is very common in creole languages (Du Toit (1905:94-95), Valkhoff (1972:23), Makhudu (1984:89), Stolz (1986:180), Holm (1988:157), Van Rensburg (1989:146-147) and Den Besten (1988:32-40)). More concretely, both Bouman (1926:40)

[21] See footnote 18 regarding the use of the term 'Coloured'.

and Valkhoff (1972:23-24) argue that this combination of two verbs (i.e. the PC verb and the lexical verb) without any intervening element such as *en* is a borrowing from the syntax of Khoisan languages.

However, as Roberge (1994:64) points out, there is no direct evidence that the combination of *loop* or one of the posture verbs plus a lexical verb is borrowed from the basilect: there simply are not enough texts from the relevant period to confirm this hypothesis. He suggests that modern day Afrikaans PC constructions are what he calls a ‘hybrid structure’, the result of a compromise (‘rapprochement’ as he labels it) between the verbal hendiadys from Middle and Early Modern Dutch (e.g. (46)–(47)) on the one hand, and the serial verb construction from the basilect (e.g. (48)–(49)) on the other. The basilectal quirky V2 construction was also ‘Netherlandicised’, by moving only the PC verb into the V2 position, and leaving *en* plus the lexical verb in clause-final position. In other words, structures like the one in (48) were ‘adapted’ to a more ‘Netherlandic’ syntax, by adding *en* in between the PC verb and the lexical verb, and in main clauses by only having the PC verb in V2 position. This resulted in the now ‘standard’ Afrikaans PC constructions for clause-final (50) and V2 environments (51).

(50) Hij *het* die koeie elke dag **staan *en* melk**.
he has the cows every day stand and milk
‘He has been (standing and) milking the cows every day.’

(51) Hij **staan** die koeie elke dag ***en* melk**.
he stand the cows every day and milk
‘He is (standing and) milking the cows every day.’

According to Roberge (1994), the compromise went both ways. That is, the European speakers of Cape Dutch did not resist the *en*-less and quirky V2 PC constructions from the basilect, while the basilectal Afrikaans speakers also took over the ‘Netherlandicised’ forms with *en* and normal V2. This is also visible today, in for example Orange River Afrikaans, where due to pressure from the standard language, more and more speakers also accept PC constructions with *en* (Roberge 1994:66).

Importantly, the andative use of *loop* was always without *en*, and furthermore, has been attested from as early the beginning of the 18^{th} century (Roberge 1994:56). An example is given in (52).

(52) Jan Dirx daar op zeijde: ***loop* begraaf** die hottentot.
Jan Dirx there on said: walk bury that hottentot
‘Jan Dirx said thereupon: go bury that Hottentot.’
(Roberge 1994:56), (source from 1707)

The andative use of *loop* thus grammaticalised much earlier than the progressive/durative use of this verb (which developed together with the posture verbs), and should be seen as an earlier, at least partly separate, development (Biberauer 2019b:15). The origins of this construction lie in imperatives

(Roberge 1994:56), (Biberauer 2019b:13). Two early uses of andative *loop* that are clearly imperative are given in (53) and (54).

(53) Gij bent beschonken, **loopt slaapen**.
you are drunk walk sleep
'You're drunk! Go (and) sleep!'
(Biberauer 2019b:13), (source from 1774)

(54) Hier is water. **loop laaf** hem.
herer is water walk revive him
'Here is water. Go (and) revive him!'
(Biberauer 2019b:13), (source from 1797)

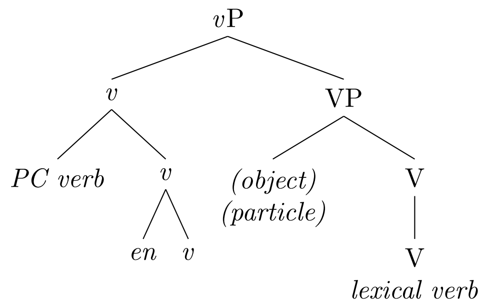
The sporadic use of *en* in PC constructions with andative *loop* should thus probably be seen as an overgeneralisation of the late 'Netherlandicised' PC construction with *en*.

Summing up, modern day Afrikaans PC constructions developed quite late, and started out from the basilectal structure without *en*, and as quirky V2 structures in non-embedded contexts, with the exception of the andative use of *loop*, which developed much earlier. In all other PC constructions *en* was introduced late, but highly successfully, in the construction in order to adapt it to a more 'Netherlandicised' syntax. The same happened with the normal V2 configurations. As the adaption was two-ways, both from the European Cape Dutch speakers and from the basilectal Afrikaans speakers, quirky V2 configurations became optional for all speakers, and *en* is sometimes left out by the European Cape Dutch speakers, and sometimes added by the basilectal Afrikaans speakers. Note, though, that even though Roberge (1994) helps us gain insight in the fact that *en* came about late in the structure, and through 'Netherlandicisation' of the basilect serial-verb construction, he does not make any claims in relation to the actual morphosyntactic status of *en*. The diachronic insights we have gained from his work makes it clear, though, that modern day Afrikaans PC constructions did not grammaticalise from real coordination constructions. This in turn means that we cannot just assume that *en* grammaticalised from coordinator into subordinator. It was introduced in the construction to 'Netherlandicise' a basilectal structure in which the PC verb and the lexical verb were structurally very close to each other. This construction then was acquired naturally by the next generation. How that generation of language learners analysed the structure, is not easy to determine. I now turn to two syntactic proposals about the structure of modern day Afrikaans PC constructions and the position of *en* in these structures.

10.6.3 De Vos (2005)

De Vos (2005), in this dissertation about English and Afrikaans pseudocoordination, is brief about the morphosyntactic status of *en*. According to him, *en* in Afrikaans PC constructions is a 'real coordinator lexeme', which is head adjoined to the little *v* of the PC construction, thus occupying a head position (De Vos 2005:161). The syntactic structure he proposes for Afrikaans PC is given in (55). In this structure, the lexical verb is in V, with in the specifier of VP a possible merge position for verbal particles and objects (De Vos 2005:159).[22] Both the PC verb and *en* are heads adjoined to *v*. As can be seen in the structure, *en* does not project. According to (De Vos 2005:96), this is how pseudocoordination differs from real coordination: in the former *en* does not project (up to what he labels an &P), whereas in the latter it does.[23]

(55) **Syntactic structure of PC construction (De Vos 2005)**



10.6.4 Biberauer (2019b)

The idea of *en* not projecting is also used by Biberauer (2019b). She proposes two possible structures for PC constructions. The first structure is the more 'traditional' PC (56), that builds directly on the Dutch constructions, in which *en* combines a verbal head (the PC verb) with the VP of the lexical verb (Biberauer 2019b:16).[24]

As can be seen in the structure, *en* does not project here either.[25] Note that the lexical verb in this structure is represented as V, but I assume this means that inside that V, there is a root of the lexical verb that has been verbalised by a verbalising head *v*.

---

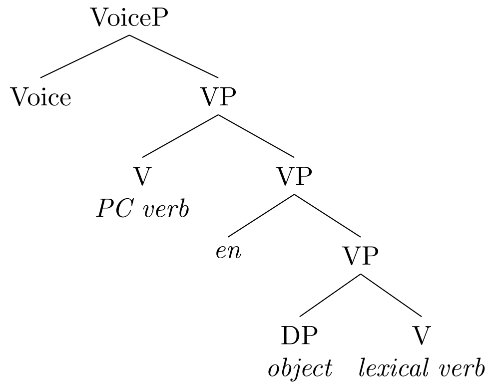
[22] Note that these particles do not occupy the same position as object, see De Vos (2005:160) for details.

[23] Note that the underlying structure that De Vos (2005) proposes for Afrikaans PC is slightly different from the one he proposes for English PC. See the original work for details.

[24] The label 'traditional' refers to the 'Netherlandicised' structure in Roberge (1994)'s terms.

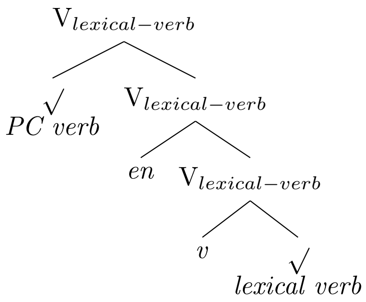
[25] Note that the position of the object in this structure represents its base-generated position, and that it scrambles out of the VP at a later point in the derivation (Biberauer 2019b:16)

(56) **'Traditional' PC construction (Biberauer 2019b)**



Alongside the 'traditional' PC construction, Biberauer (2019b) proposes an 'innovative' PC construction available in modern day Afrikaans.[26] The syntactic structure of this 'innovative' PC construction is given in in (57).

(57) **'Innovative' PC construction (Biberauer 2019b)**



In this structure, the PC verb and the lexical verb are syntactically much closer to each other than in the 'traditional' PC construction. In the 'innovative' structure, the PC verb is a root that is part of the extended V projection of the lexical verb. Like in the 'traditional' structure, *en* is positioned very low in the structure, and does not project. (Biberauer 2019b:17) herself calls *en* in Afrikaans PC constructions "acategorial", which she takes to mean that it cannot project or supply a label. Note that both structures proposed by Biberauer (2019b) for Afrikaans PC constructions are monoclausal, and in both cases the semi-lexically used verb is in the extended projection of the lexical verb. In the 'traditional' structure, the PC verb is a verbal head that combines with the VP of the lexical verb, and in the 'innovative' structure, the PC verb is a root the combines with the verbal head of the lexical verb. In both structures,

[26]Note that it is 'innovative' from a European/West-Germanic perspective; as we have seen earlier in this section, diachronically this structure was used before the 'traditional' PC construction emerged.

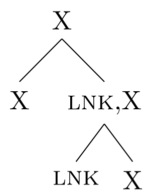
en does not project, but is merely part of the verbal extended projection.

10.6.5 *En* in the current analysis

I follow De Vos (2005) and Biberauer (2019b) in assuming that *en* is a head in the syntactic structure that does not project. The fact that *en* is unstressed and very often reduced to a schwa, and that it can be seen as an enclitic element that does not bare any semantic features, show that *en* is a very 'light' morpheme. I follow Biberauer (2019b) in proposing that it is part of the extended projection of the lexical verb, and that it is merged without changing the label of that projection. I am aware that this does not answer the question of what *en* is exactly: if it is a functional head, we would expect it to project, and if it is a root, we would expect it to have semantic features. The fact that it fits neither description is reminiscent of a class of functional heads that lacks semantic features, labeled *linkers* as in Philip (2012:5). According to her definition of linkers, the only role of these elements is to mark a grammatical relationship (like subordination or coordination), which exists independently of the presence of the linker. In other words, the linker is a mere indicator of the grammatical relation between two elements, and it can be present or absent in the same grammatical relation in different languages. According to Philip (2012) coordinators fall in the class of linkers, which means that *en* in Afrikaans PC constructions could potentially be a linker, if we would assume with De Vos (2005) that *en* in these constructions is a coordinator.[27]

Philip (2012) proposes that linkers are functional heads that project in syntax, but only once, and furthermore they do not contribute any semantic or formal features to the structure.[28] An abstract example of a coordination construction involving two syntactic heads of the same category (X) as proposed by Philip is given in (58) (Philip 2012:127).

(58) **Coordination of heads (Philip 2012)**



As can be seen in this structure, the linker structurally intervenes between the two heads. According to Philip (2012) this holds for all linkers that are independent heads in the syntax. What we can also see in the structure is that the linker projects, but only once. Philip (2012) argues that a linker can only

[27] The idea that coordinators are linkers has also been proposed by Dik (1983) and Zwart (2009:1599).

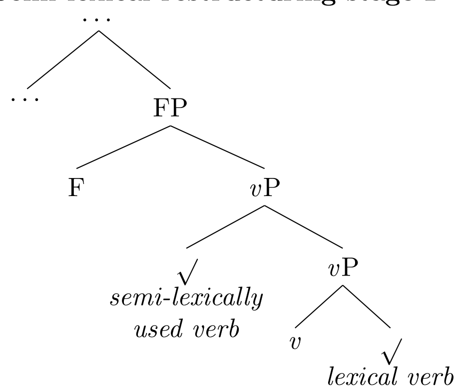
[28] The view that coordinators are semantically vacuous can also be found in Blümel (1914:52), Hockett (1958:153) and Zoerner (1999:232).

project once because it does add any features. The head X itself projects as well. However, if a linker projects only once because it contains neither syntactic nor semantic features, we might as well say that linkers do not project at all. I therefore propose that *en* in Afrikaans PC constructions should indeed be seen as a linker, but that it does not project. Of course this still does not solve the problem of what this element is, except that it is a linker without any syntactic or semantic features, which is not a root. Unfortunately, solving this problem is beyond the scope of this dissertation, and needs to be left for future research. For now, I just assume that there can be such a thing as a non-projecting non-root in human language (see Weisser (2015), Biberauer (2017b), Song (2019) for discussion on why these types of elements quite likely form a natural class in language.)

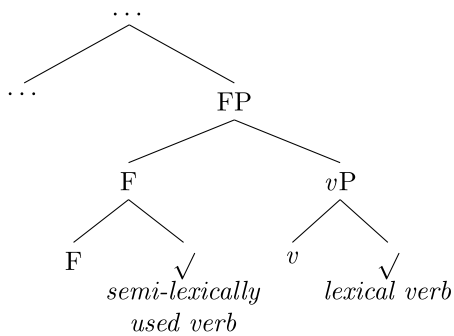
10.7 The structure of the PC constructions

In this section, I present the structures for the Afrikaans PC constructions. As we have seen in chapter 2 and case study I, there are two stages of semi-lexicality in the verbal domain, which I assume is also the case for Afrikaans. The general structures proposed in this thesis for the two stages of semi-lexicality are repeated here in (59) and (60).

(59) **Semi-lexical restructuring stage I**

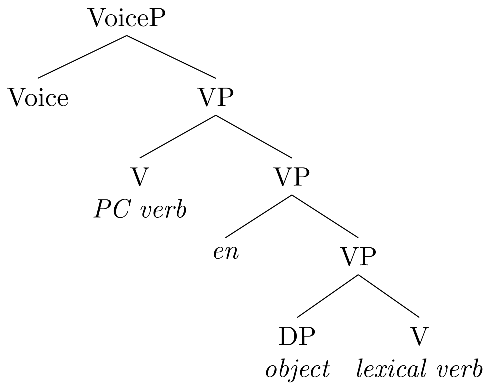


(60) **Semi-lexical restructuring stage II**

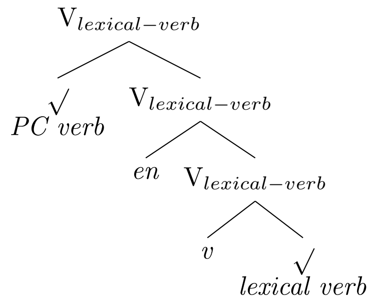


However, I need to make one adaption regarding the first stage of semi-lexicality which holds specifically for Afrikaans syntax (and which is hence not a valid structure in Dutch). Recall from section 10.6 that Biberauer (2019b) proposes two possible underlying structures for Afrikaans PC constructions, a 'traditional' one, modelled on Dutch syntax and probably the result of standardisation and normative pressure, and an 'innovative' one, which is most probably due to intense language contact with the basilect, in which PC verbs could be used as prefixes on the lexical verb. While Biberauer (2019b) does not make a distinction between a first and second stage of semi-lexicality of Afrikaans PC verbs per se, I want to propose that her two underlying structures are two possible structures for the first stage of semi-lexicality. I repeat the two structures as proposed by Biberauer (2019b) here in (61) and (62).

(61) **'Traditional' PC construction (Biberauer 2019b)**

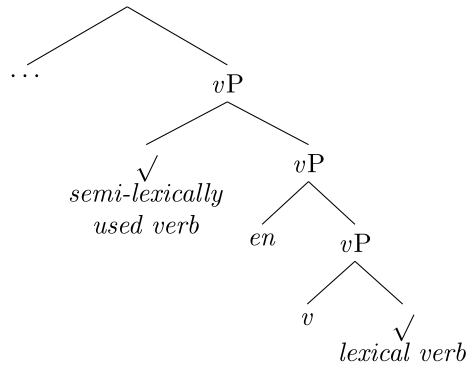


(62) **'Innovative' PC construction (Biberauer 2019b)**



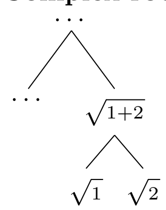
The main difference between the two structures is that in the 'traditional' structure the semi-lexical and the lexical root are structurally less close to each other than in the 'innovative' structure. I also adopt Biberauer (2019b)'s idea that in the 'innovative' structure, the semi-lexical and the lexical root are structurally much closer than in the 'traditional' structure. Adapting this idea to the structure of the first stage of semi-lexicality proposed in this thesis, I propose that the structure for stage I, given in (63) is the 'traditional' structure. Specific to Afrikaans syntax, *en* is inserted in the functional structure above the lexical root. Recall from section 10.6 that *en* is taken to be a linker that does not project in syntax. Its presence in the structure is due to standardisation and 'Netherlandicisation' of the structure (see section 10.6). This means that *en* is only expected to be present at the first stage of semi-lexicality of PC verbs, and not at the first stage of semi-lexicality of other Afrikaans semi-lexically used verbs. The 'traditional' Afrikaans structure of PC constructions adopted in this thesis is given here in (63).

(63) **The 'traditional' semi-lexical restructuring stage I**



Recall that I adopt Biberauer (2019b)'s idea that the semi-lexical and the lexical root are structurally much closer to each other in the 'innovative' structure than in the 'traditional' one. In the structure that I propose, the two roots are directly Merged with each other at the beginning of the derivation. The idea that two roots should be able to be combined without any intervening elements has already been proposed before, for example by Zhang (2007) for Chinese compounds, and for derivational suffixes and lexical roots by Lowenstamm (2014). When two roots are directly Merged with each other, the problem that arises is how this root complex gets labeled.[29] Zhang (2007:177-178) solves this by proposing that the decision for the category of the root complex is postponed until a functional categorising head is Merged. For her, complex roots have the structure given in (64), in which '$\sqrt{1+2}$' should not be seen as an actual syntactic label. The label will be given by the verbalising head that will be Merged with the complex root (Zhang 2007:177).

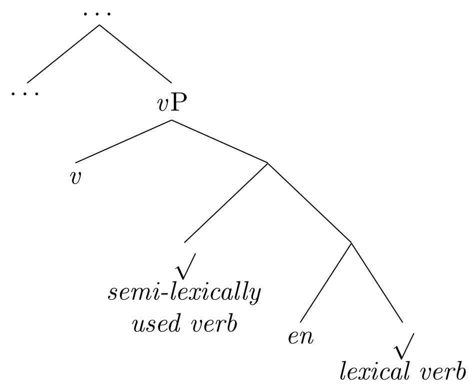
(64) **Complex roots (Zhang 2007)**



Similarly, Rizzi (2016) argues that labeling can be delayed until the end of the phase, meaning that the syntactic derivation can go on even when a node remains without a label as long as the construction of a given phase is not yet finished. In contrast, Lowenstamm (2014) assumes that roots can project as well, which means that the label of two Merged roots is $\sqrt{}$P. Here, I follow Zhang (2007) and Rizzi (2016), and assume that two roots can be Merged without being labeled. The label of the complex root is given by the verbalising head that is Merged with the root complex. Furthermore, specifically for the Afrikaans PC constructions, I assume that the linker *en* is part of the complex root as well, and that it structurally intervenes between the two roots (Philip 2012). The structure of the 'innovative' PC constructions is given in (65).

[29] Interestingly, Zhang (2007) shows that in Chinese, compounds can have two underlying structures, one in which the two roots are Merged directly and can therefore never be separated by movement, and one in which the two roots have their own functional material, and can be separated by movement. This looks very similar to normal vs. quirky V2 in Afrikaans PC constructions, which, as I propose below, is also explained by the two underlying structures.

(65) **The ‘innovative’ semi-lexical restructuring stage I**

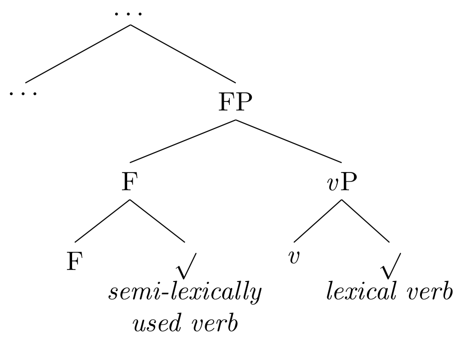


I assume that both the ‘traditional’ and the ‘innovative’ structure for Afrikaans PC constructions at stage I are part of the grammar of most Afrikaans speakers – more specifically, of those that have received a formal education in Afrikaans. In other words, the fact that both structures exist for these speakers is due to normative instruction in school for the ‘traditional’ structure.

The second stage of semi-lexicality generally proposed in this thesis remains the same for Afrikaans, and is repeated below in (66). In this structure, the semi-lexically used root is Merged in a separate workspace with a functional head. This complex head is Merged in the functional projection of the lexical root.

Summing up, I have proposed that the first stage of semi-lexicality in the Afrikaans verbal domain has two possible underlying structures: the ‘traditional’ one, in which the verbalising head separates the semi-lexical and the lexical root, and the ‘innovative’ one, in which the verbalising head is structurally above a complex root. The second stage of semi-lexicality has only one possible underlying structure, in which the semi-lexically used root forms a complex head with a functional head above the lexical root. With these structures in place, we move on to the analysis of the specific PC constructions discussed in this case study.

(66) **Semi-lexical restructuring stage II**



10.8 The analysis of the configurations per PC verb

In this section, I present the analysis of the PC configurations per PC verb, in the following order: $loop_{andative}$, $loop_{progressive}$, *sit*, *staan* and *lê*. For each PC verb, I illustrate how the structures proposed in the previous subsection explain the morphoysyntactic variation and optionality in the PC constructions as presented in the previous chapter, namely the presence/absence of *en*, the presence/absence of *ge-*, and normal versus quirky V2.

10.8.1 PC verb *loop*$_{andative}$

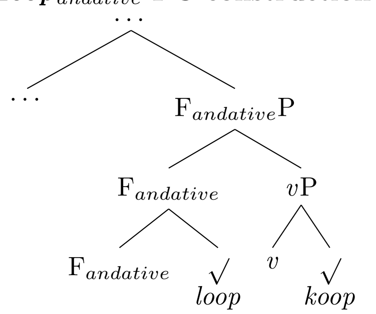
I start with the analysis of the morphosyntactic variation and optionality in PC constructions with $loop_{andative}$. Recall from subsection 9.2.1 that this verb was only tested in embedded PC constructions, and not in V2 PC constructions. The test item that was used to investigate the presence/absence of *en* and that of *ge-* in this cluster is given in (67).

(67) Paul sê dat Lisa verlede week 'n splinternuwe motor **(ge)loop**
Paul says that Lisa last week a splinter.new car GE-walk
(en) koop het.
and buy has
'Paul says Lisa went and bought a brand new car last week.'

The main findings for this PC verb were that (i) it has the most distinct morphosyntactic behaviour compared to the other four verbs, (ii) *en* is most frequently absent, but its presence is not very infrequent either, (iii) *ge-* is most frequently absent, but its presence is very frequent as well.

Recall that in section 10.3, I established that $loop_{andative}$ is at the second stage of semi-lexicality. Based on the general structures proposed in the previous section, the structure of the verb cluster with $loop_{andative}$ that was tested in the questionnaire study is given in (68).

(68) ***Loop****andative* **PC construction**



In the derivation of this PC construction, the semi-lexically used root *loop* is Merged in a separate workspace with a functional *andative* head. This complex head is Merged in the functional structure of the lexical root *koop* 'buy', after this root has been verbalised by *v*. Note that the verb *gaan* 'go' in Afrikaans can also be used to express an andative reading (69). Moreover, given that it can embed PC verbs (also shown in (69)), it is at the second stage of semi-lexicality as well. The fact that this use of *gaan* is highly frequent probably serves as a cue for the language learner that a feature such as $\mathrm{F}_{andative}$ exists in the language.

(69) Die studente het ***gaan*** **sit en** notas **maak**.
the students have go sit and notes make
'The students just went to sit and make notes (after a lecture).' (Biberauer 2019b:11)

The fact that *loop*$_{andative}$ is at the second stage of semi-lexicality while all other PC verbs are in the first stage (see section 10.3), combined with the fact that I propose different structures for the two stages (see section 10.7), provides an account for the first main finding regarding the PC verb *loop*$_{andative}$, namely the fact that it has the most distinct morphosyntactic behaviour compared to the other PC verbs.

Let us now turn to the finding that *en* is mostly absent with *loop*$_{andative}$, even though occurrences with *en* are not infrequent. In the structure of the second stage of semi-lexicality, there is no *en* in the functional structure of the lexical root. We would thus expect *en* to always be absent in *loop*$_{andative}$ PC constructions. The question is thus what to make of the cases in which speakers do allow *en* in these constructions. A possible scenario is that in the previous century, when the modern day Afrikaans PC construction was formed using both the Dutch model of PC constructions with *en* and the basilect PC constructions without *en*, some Afrikaans speakers have overgeneralised the normative rule of *en*-insertion to the andative use of *loop*. Of course, such a statement can only be confirmed by an extensive diachronic corpus study on the rise of *en* in Afrikaans PC constructions, combined with an investigation into the normalisation processes concerning *en* in these constructions. Unfortunately, this is beyond the scope of this dissertation. It would clearly be a very fruitful direction for future research, though, to concentrate on the diachronic spread of *en* in PC constructions in general, and the role of standardisation and normative pressure in this process.

The third main finding for $loop_{andative}$ PC constructions is that *ge-* is most frequently absent in this construction, but that it can also quite frequently occur. In section 10.5 I have proposed that *ge-* is a past tense marker that can attach to phrasal material. In the structure in (68), it can thus attach to the complex $F_{andative}+\surd$ head. In section 10.5 I have furthermore pointed out that *ge-* is generally optional on the V2+V3 complex in three-verb clusters, due to the phonological rule which renders *ge-* optional on verbs or verbal complexes with a unstressed-stressed pattern, as is the case in PC constructions (e.g. *loop kóóp*). The optionality of *ge-* in $loop_{andative}$ PC constructions is thus exactly what we would expect. The fact that the frequency for this construction without *ge-* is slightly higher than that with *ge-* might be due to analogy with the other andative verb in Afrikaans, namely *gaan* 'go'. Recall from 10.5 that Afrikaans *gaan* is one of the Afrikaans clustering verbs that were directly inherited as IPP verbs from Dutch, and as a result, always occurs without *ge-*. Given that *gaan* can be used semi-lexically as an andative verb,[30] it is not unlikely that some Afrikaans speakers overgeneralise the *ge*-less use of semi-lexical *gaan* to the semi-lexical use of $loop_{andative}$.

As already mentioned at the beginning of this subsection, V2 PC constructions with $loop_{andative}$ were not tested. According to Biberauer (2017c:11) V2 PC constructions with *loop* in which *en* is absent, obligatorily occur in a quirky V2 configuration, as illustrated in the following example.

(70) a. ***Loop*** **koop** hy toe vir hom 'n nuwe kar?
walk buy he then for him a new car
'Did he then go and buy himself a new car?'

b. ****Loop*** hy toe vir hom 'n nuwe kar **koop**?
walk he then him a new car buy

(Biberauer 2017c:11)

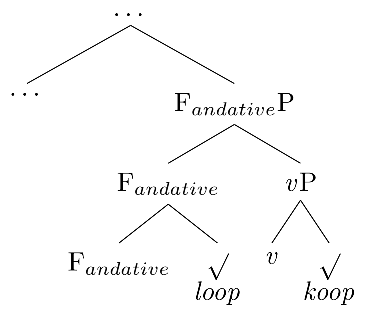
I want to propose the following for the fact that $loop_{andative}$ always occurs in the quirky V2 configurations. Assume, for the sake of the argument, that V2 movement happens in narrow syntax.[31] In the proposed syntactic structure for PC constructions with $loop_{andative}$, repeated here in (71), we can observe a symmetric relation between the complex head $F_{andative}$, which consists of $F_{andative}$ and the semi-lexically used root, and the lexical *v*P, which consists of

[30] It can also be used as a future auxiliary (De Vos 2001).

[31] I am aware of the fact that V2 has also been analysed as a PF phenomenon, for example by (Chomsky 2001:137); see Holmberg (2015) for discussion of syntactic versus PF analyses of V2.

v and the lexical root.

(71) ***Loop**$_{andative}$* **PC construction**



My proposal is that this symmetric structure results in two options for the size of the element that is being moved to V2. The first option is that only the complex F$_{andative}$ head is moved to V2 position, resulting in normal V2. The second option is that the element to be moved is selected one level up, namely the entire F$_{andative}$P, which includes the lexical verb. This would result in quirky V2. The question now is why quirky V2 is the only possible V2 configuration with *loop*$_{andative}$. Again, I hypothesise that this is a side-effect of the standardisation of the PC constructions, which did not happen with other Afrikaans semi-lexically used verbs. Recall from section 10.6 that for all PC verbs except *loop*$_{andative}$, the original quirky V2 construction from the basilect was 'Netherlandicised'. This 'Netherlandicisation' involved two adaptations: *en* was added between the PC verb and the lexical verb, and the quirky V2 constructions were changed into 'normal' V2 structures in which *en* and the lexical verb remained in situ. However, the quirky V2 constructions were also adopted into the 'standard' Afrikaans syntax as well, given that it was a two-way adaptation between the basilect and 'European' Afrikaans. The result was an *en*-less quirky V2 construction, and a normal V2 construction with *en*. As I have already speculated above, it is quite likely that the andative use of *loop* was influenced by these processes, and by later normative pressure at schools to keep the 'Netherlandicised' structure with *en* intact. The fact that for the other PC constructions, the quirky V2 configuration later also contained *en*, can then be seen as a generalisation by the next generation of language learners acquiring the 'standard' Afrikaans variety that all PC constructions expressing progressive/durative aspect contain *en*, including in their quirky V2 configurations. For *loop*$_{andative}$, this generalisation was not made, because this verb usually occurs without *en*. In other words, based on the syntactic structure of *loop*$_{andative}$ constructions ((71)), I would predict that both normal and quirky V2 movement is possible. However, the normal V2 configuration is ruled out due to a generalisation by language learners after the two-way adaptation between 'Netherlandicised' and basilect PC constructions, that PC constructions without *en* always have to be a quirky V2 configuration. This generalisation was not made for other Afrikaans semi-lexically used verbs, such as *gaan* 'go', *laat* 'let', or *kom* 'come', because they were never associated with the subclass of PC verbs by language learners. All these verbs are thus expected to show optional quirky V2, which is indeed the case (De Vos 2001, Biberauer 2019b, Augustinus and Dirix 2019).

10.8.2 PC verb *loop*$_{progressive}$

Let us now move on to the PC verb *loop*$_{progressive}$. The test item that was used to test this verb in embedded configurations is repeated in (72), and the test item that was used to test this verb in V2 configurations is repeated in (73).

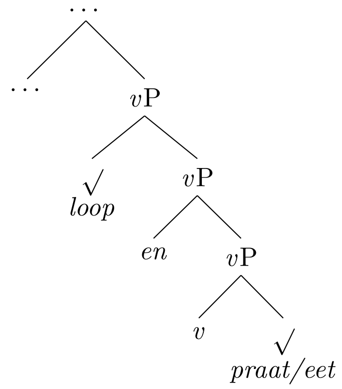
(72) Steve sê dat Cornelia gisteraand baie **(ge)loop (en) praat** het.
Steve says that Cornelia yesterday.night a.lot GE.walk and talk has
'Steve says that Cornelia was (walking and) talking a lot yesterday night.'

(73) Hoekom **loop** $<$ **(en) eet** $>$ Jan heeldag piesangs $<$ **en eet** $>$?
Why walk and eat Jan all.day bananas and eat
'Why is Jan (walking and) eating bananas all day long?'

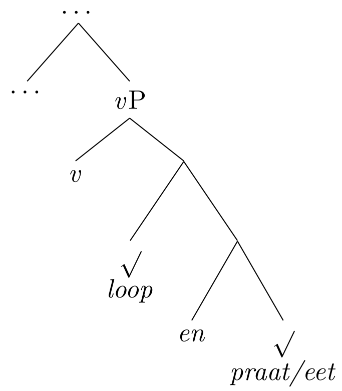
The main findings for this PC verb were that (i) its behaviour is slightly different from the other PC verbs, (ii) *en* is almost always present, though it can be absent for a limited number of speakers as well, (iii) *en*-absence is more frequent in quirky V2 constructions than in all other constructions, (iv) *ge-* is optional, with slightly higher frequencies for it being present, and (v) normal and quirky V2 constructions are optional, with slightly higher frequencies for quirky V2.

Recall from section 10.3 that *loop*$_{prog}$ is at the first stage of semi-lexicality. In the previous section, I have proposed that the first stage of semi-lexicality has two possible underlying structures, a 'traditional' one, modelled on Dutch, and an 'innovative one, modelled on the basilect. The underlying structures for the PC constructions with *loop*$_{prog}$ as tested in the questionnaire study thus have the following two possible underlying structures.

(74) ***Loop***$_{progressive}$ **'traditional' structure**



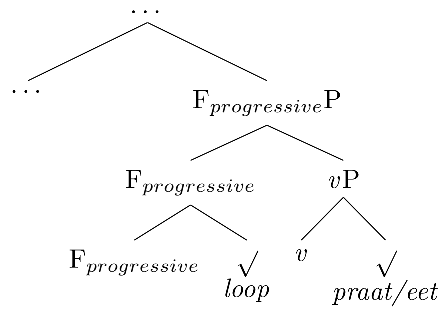
(75) ***Loop***$_{progressive}$ **'innovative' structure**



As can be seen from the structures, *en* is present in both. We would therefore expect *en* to always be present, whereas the data have shown that *en* is allowed to be absent by a small number of speakers. Recall from subsection 9.3.6, though, that *loop*$_{progressive}$ is semantically quite bleached, more so than the PC verbs *sit* and *lê*. This might be an indication that for some speakers *loop*$_{progressive}$ is grammaticalising from the first stage of semi-lexicality onto the second one, given that grammaticalisation often results in further semantic bleaching (Hopper and Traugott 1993). For those speakers, the structure of PC constructions with *loop*$_{progressive}$ is the same as the one for *loop*$_{andative}$, given here in (76).[32]

[32]Note that the label 'progressive' on the functional head in this structure should not be seen as a very specific progressive feature, but as a higher level, aspect related feature which includes progressive (cf. Ramchand (2018:72) on the core property of the progressive being an 'Identifying state'.)

(76) ***Loop***$_{progressive}$ **PC construction - stage II**



In this structure, *en* is absent. Note also that *en* is most often absent in PC constructions with *loop*$_{progressive}$ in quirky V2 configurations. In the previous subsection, we have seen that *loop*$_{andative}$ always occurs in quirky V2 configurations in V2 contexts. If for some speakers *loop*$_{progressive}$ is at the second stage of semi-lexicality, and in this stage of PC constructions quirky V2 is the obligatory V2 configurations, we thus indeed expect that the highest frequency of *en* absence lies in quirky V2 configurations. Given that there are no clear indications that *loop*$_{progressive}$ is grammaticalising into the second stage of semi-lexicality in the language area as a whole, *en* absence with *loop*$_{progressive}$ is expected to be quite infrequent, which is indeed the case.

One of the other main findings can also be related to the grammaticalisation of *loop*$_{progressive}$ into the second stage of semi-lexicality for some speakers, namely the fact that the frequency of quirky V2 is slightly higher than that of normal V2. For the small group of speakers for whom this verb is already in the second stage of semi-lexicality, the only possible configuration is quirky V2. However, for the majority of the speakers, *loop*$_{progressive}$ is at the first stage of semi-lexicality, and has thus two possible underlying structures in PC constructions, the ones given in (74) and (75). I follow Biberauer (2019b) in proposing that the 'traditional' structure gives rise to normal V2 configurations, whereas the 'innovative' structure results in quirky V2. In the 'traditional' structure, the coordination is asymmetric in nature: a root is being coordinated with a *v*P (see also Biberauer (2019b) on the asymmetric nature of 'traditional' PC constructions). Because of this asymmetry, the entire PC construction is not a target for V2 movement, only the highest verb is, i.e. the PC verb. Thus, the 'traditional' structure results in normal V2. In the 'innovative' structure, the coordination is symmetric, however: two roots are being coordinated, and together they form a complex root. They are therefore seen as a single target for V2 movement, and so the 'innovative' structure results in quirky V2. This means that for the larger group of speakers for whom *loop*$_{progressive}$ is at the first stage of semi-lexicality, normal/quirky V2 is truly optional. However, as there is also a small group of speakers for whom *loop*$_{progressive}$ is at the second stage of semi-lexicality and for whom quirky V2 is obligatory, the total frequency distribution among all speakers is therefore expected to tilt slightly towards quirky V2, which is indeed one of the main findings for this PC verb. The fact that for some speakers this verb is at the second stage of semi-lexicality also explains why this verb has a slightly different profile compared to *sit* and *lê*, which show much less semantic bleaching. Furthermore, *staan* behaves slightly more like $loop_{progressive}$ than the other two posture verbs, and *staan* is also more semantically bleached than the others.

The last main finding that needs to be explained is the fact that *ge-* is optional in the embedded PC constructions, even though it is slightly more frequently present than it is absent. Given that *ge-* can attach to phrasal material, it should be able to attach to the *v*P level of all structures in which $loop_{progressive}$ can occur: the 'traditional' structure of stage I, the 'innovative' structure of stage I, and the structure of stage II. As I have already explained in the previous subsection, *ge-* is generally optional due to the stress pattern of the cluster being unstressed-stressed. The fact that *ge*-presence is more frequent could possibly be a subconscious strategy of Afrikaans speakers to sharpen the distinction between $loop_{andative}$ – which has higher frequencies for *ge-* being absent – and $loop_{progressive}$. This remains a hypothesis, however, as I do not have any evidence for this scenario, and it would require sociolinguistic research in order to be tested.

10.8.3 PC verb *sit*

We now turn to *sit*. Embedded PC constructions with this verb were tested with the test items given in (77). The V2 constructions with *sit* were tested with the test item given in (78).

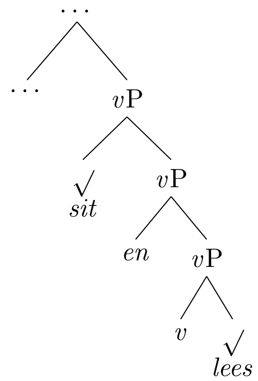
(77) Simon sê dat Thomas die hele middag **(ge)sit (en) lees** het.
Simon says that Thomas the entire afternoon GE.sit and read has
'Simon says Thomas has been (sitting and) reading the entire afternoon.'

(78) Hoekom **sit** < **(en) lees** > Lisa heeldag die koerant < **(en) lees** >?
Why sit and read Lisa all.day the newspaper and read
'Why is Lisa (sitting and) reading the newspaper all day long?'

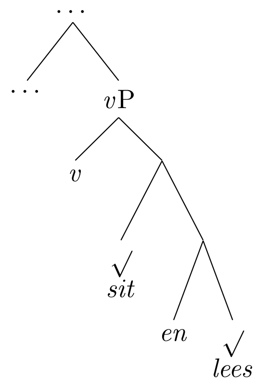
The main findings for this PC verb were that (i) *en* is always present, (ii) *ge-* is truly optional, and (iii) normal vs. quirky V2 is truly optional.

Recall from section 10.3 that *sit* is at the first stage of semi-lexicality. The two possible underlying structures for the PC constructions with this verb are thus as follows.

(79) ***Sit* 'traditional' structure**



(80) **$\textit{\textbf{Sit}}_{progressive}$ 'innovative' structure**



Recall from subsection 9.3.6 that *sit* is not semantically bleached at all, which means that we do not have any indication for this verb being in the process of grammaticalising from the first stage of semi-lexicality to the second. Given the very uniform behaviour of *sit* for all Afrikaans speakers, the analysis of PC constructions with this verb is quite straightforward. The fact that *en* is always present is the result of *sit* being in the first stage of semi-lexicality for all speakers: in both structures of the first stage, *en* is part of the PC construction. Regarding the true optionality of *ge*-, we can say the following. As already explained in the previous two subsections, *ge*- can attach to vP, and is thus expected to occur in embedded PC constructions. Its optionality is caused by other factors independently of the PC configuration per se. The fact that *ge*- is truly optional is thus exactly what we would expect for PC constructions with *sit*. The same holds for the optionality of normal vs. quirky V2. As explained in the previous subsection, the 'traditional' structure results in normal V2, whereas the 'innovative' one results in quirky V2. Since both structures are possible structures in the first stage of semi-lexicality, true optionality of normal vs. quirky V2 is indeed expected for the PC verb *sit*.

10.8.4 PC verb *staan*

Let us now look at the PC verb *staan*. The embedded PC constructions with this verb were tested with the test item in (81). The V2 constructions with this verb were tested with the test item in (82).

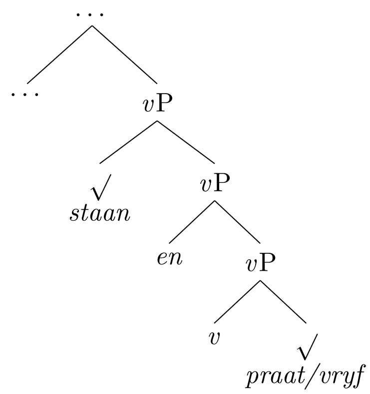
(81) Susan sê dat Elsa vir ure met haar ma op die telefoon
Susan says that Elsa for hours with her mom at the telephone
(ge)staan (en) praat het.
GE.stand and talk has
'Susan says that Elsa has been (standing and) talking with her mom on the phone for hours.'

(82) Hoekom **staan** < **(en) vryf** > Thomas heeldag sy ken < **(en)**
Why stand and rub Thomas all.day his chin and
vryf >?
rub
'Why is Thomas (standing and) rubbing his chin all day long?'

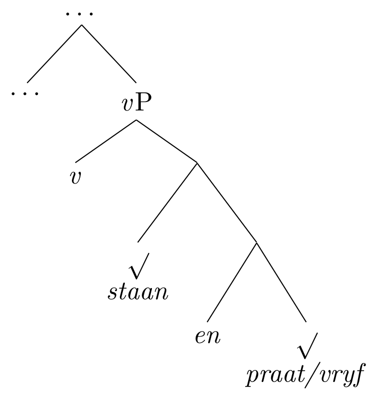
The main findings for PC constructions with this verb were that (i) its behaviour is slightly different from the other two posture verbs, *sit* and *lê*, (ii) *en* is almost always present, but can be absent for a small number of speakers, (iii) *ge-* is truly optional in embedded PC constructions, and (iv) normal vs. quirky V2 is truly optional.

Recall from section 10.3 that *staan* is at the first stage of semi-lexicality. This means that PC constructions with *staan* as tested in the questionnaire study have two possible underlying structures, given in (83) and (84).

(83) ***Staan* 'traditional' structure**

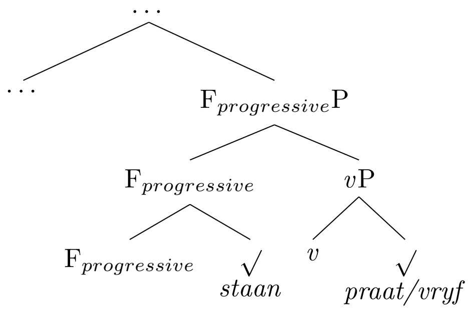


(84) ***Staan* ‘innovative’ structure**



In both structures, *en* is present. This explains the high frequency of *en*-presence, but not the low frequency of *en*-absence. As I have proposed for $loop_{progressive}$, it seems that for some speakers the PC verb *staan* is grammaticalising from the first stage of semi-lexicality onto the second. An indication for this is the fact that *staan* is semantically quite bleached. This means that for some speakers, the structure of *staan* PC construction is as in (85).

(85) ***Staan* PC construction - stage II**



In this structure, *en* is absent, and given that for a small group of speakers this is the underlying structure for *staan* PC constructions, low frequencies of *en*-absence are thus expected. The fact that *staan* is at the second stage of semi-lexicality for some speakers explains why it behaves slightly differently from the other two posture PC verbs, *sit* and *lê*, which are in the first stage for all speakers. Note furthermore that *staan* has reached the second stage of semi-lexicality for a smaller group of speakers than $loop_{progressive}$: in subsection 9.3.9, I have established the following hierarchy for the five PC verbs:

(86) $loop_{andative} > loop_{progressive} > staan > lê/sit$

The fact that the PC verb *staan* behaves morphosyntactically ‘in between’ *loop*$_{progressive}$ on the one hand and the two posture verbs on the other is reflected by the fact that it shows low frequencies of *en*-absence, like *loop*$_{progressive}$, but true optionality of *ge-* and normal vs. quirky V2, like *lê* and *sit*. The explanation of these latter two main findings is identical to the one I have given in the previous subsection for the PC verb *sit*. That is, *ge-* can in principle attach to *v*P in both underlying structures of stage I, but it is rendered optional by other factors. Normal V2 is the result of V2 movement starting from the 'traditional' structure, while quirky V2 is the result of V2 movement starting from the 'innovative' structure. The small group of speakers that also has the underlying structure of stage II for the PC verb *staan* has apparently not affected the frequency distribution of normal versus quirky V2.

10.8.5 PC verb *lê*

The last PC verb in need of an analysis is *lê*. The embedded PC constructions with this verb were tested with the test item given in (87), and the V2 constructions with the test item given in (88).

(87) Eric sê dat Michael die hele naweek **(ge)lê (en) slaap** het.
Eric says that Michael the entire weekend GE.lie and sleep has
'Eric says that Michael has been (lying and) sleeping the entire weekend.'

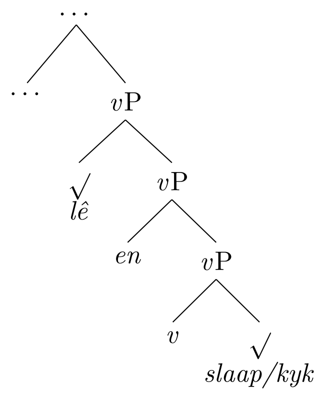
(88) Hoekom **lê** < **en kyk** > Mark heeldag na die wolke < **en kyk**> ?
Why lie and look Mark all.day at the clouds and look
'Why is Mark (lying and) looking at the clouds all day?' .

The main findings for PC constructions with this verb were that (i) *en* is almost always present, with very low frequencies of *en* absence, (ii) *ge-* is truly optional in embedded PC constructions, and (iii) normal vs. quirky V2 is truly optional.

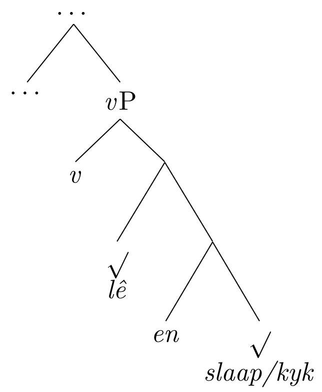
Recall from section 10.3 that the PC verb *lê* is at the first stage of semi-lexicality. This means that the two possible underlying structures for the PC constructions with this verb are given in (89) and (90).

Given that *lê* is hardly semantically bleached at all (see subsection 9.3.9), there is no evidence for this verb being in the process of grammaticalising from stage I to stage II of semi-lexicality. It is therefore unexpected that there are a number of speakers who allow *en* to be absent in PC constructions with *lê*. Note, though, that the frequencies for *en*-absence with *lê* are much lower than the – already low – frequencies of *en*-absence with *loop*$_{progressive}$ and *staan*. In fact, the frequencies of *en* absence in the configurations with *lê* that are above the 5% noise threshold, are only just above this threshold: 5,1% for the embedded PC configuration without *ge-*, and 5,7% for the quirky V2 configuration. For now, the only explanation I can provide for *en* absence with *lê* is that it is noise that just exceeded the set threshold of 5%. A replication of the questionnaire study could serve as a testing ground for this hypothesis in future research.

(89) ***Lê* ‘traditional’ structure**



(90) ***Lê* ‘innovative’ structure**



The optionality of *ge*- and normal vs. quirky V2 can be explained in the same way as I have done in the previous subsections for *sit* and *staan*. *Ge*- optionality is the result of *ge*- being optional on V2+V3 in three-verb clusters due to a phonological condition. Both normal and quirky V2 are possible because of the double underlying structure of the first stage of semi-lexicality.

CHAPTER 11

Conclusion and outlook

This case study has shown that in the Afrikaans verbal domain it is necessary to make a distinction between two stages of semi-lexicality, as was the case in the first case study for Dutch. Based on the empirical findings of this case study, I have been able to confirm the following morphosyntactic scale for the five PC verbs in Afrikaans (see De Vos (2001) and Biberauer (2019b)):

(1) $loop_{andative} > loop_{progressive} >$ *staan* $>$ *lê*/*sit*

$Loop_{andative}$ is at the second stage of semi-lexicality for all speakers, whereas on the other side of the spectrum, *lê* and *sit* are uniformly in the first stage. For most speakers, $loop_{progressive}$ is at the first stage of semi-lexicality, but for a number of them this verb is grammaticalising into the second stage of semi-lexicality. The same holds for *staan*, though the grammaticalisation into the second stage is happening with even less speakers than is the case for $loop_{progressive}$.

In subsection 9.3.7, the exploratory statistical techniques suggested that two of the main optionality patterns, namely the optionality of *en* and that of *ge*-, should be seen as two separate phenomena, and furthermore that the presence or absence of *ge*- occurs in both 'standard' and 'non-standard' Afrikaans. The analysis has shown that the presence or absence of *en* and *ge*- are indeed completely separate phenomena. The optionality of *ge*- in Afrikaans in general was explained by it being able to attach to phrasal material, it being an optional additional past tense marker, and it being phonologically optional based on the stress pattern of PC constructions, all factors that hold both for 'standard' and 'non-standard' Afrikaans.

The third optionality pattern, normal vs. quirky V2, was shown to be a side effect of the complex diachronic development of Afrikaans PC constructions, in which basilectal PC constructions were 'Netherlandicised', while the basilectal quirky V2 constructions were adopted by 'European' Afrikaans speakers in turn. This resulted in two possible underlying structures for the first stage of semi-lexicality of PC constructions, a 'traditional' one in which the semi-lexical and the lexical root are structurally separated by a verbalising head, and an 'innovative' one, in which the two roots are directly Merged with each other, forming a complex root. The presence of *en* in the progressive/durative PC constructions was also the result of 'Netherlandicising' the basilect PC constructions, which possibly resulted in *en* 'leaking' into the constructions with $loop_{andative}$ to a lesser extent as well. As the use of *en* with progressive/durative PC constructions is still actively taught in schools, normative pressure is keeping the 'Netherlandicised' PC construction the dominant one in the regions where 'European' Afrikaans is spoken, and it is slowly also being adopted by Afrikaans from other regions, like Orange River Afrikaans. Unfortunately, the questionnaire of this case study did not reach speakers of that variety, which can hopefully be achieved in future research.

Specific to this case study is that we have seen that language contact and standardisation/normative pressure has had and still has a strong influence on the morphosyntactic structures of Afrikaans PC constructions, and the optionality of *en*, *ge-* and normal vs. quirky V2 in these constructions. Not unexpectedly, these constructions are a sort of 'compromise' between EWG – more specifically, Dutch syntax – and the basilect. A future deeper diachronic investigation plus sociolinguistic research could shed more light on this.

CHAPTER 12

Discussion and general conclusion

> *'n Mens begin so rondom veertien dink. Vóór dit dink iets anders vir jou. Miskien dieselfde iets wat vir die diere dink? Miskien dink jy voor jy veertien is sonder dink. Hy weet niet. Hy weet net dat die ander dink op 'n dag in jou begin. Van binne af. Ook nie altyd nie. Soms dink jy jouself in 'n bosbraam vas omdat jy nie na die dink wat van binne af kom, luister nie.*
>
> - Kringe in 'n bos, *Dalene Matthee*

12.1 Introduction

In this final chapter, I focus on two topics that have been a common thread throughout the thesis, namely grammaticalisation and (morpho)syntactic optionality. I discuss these topics from a more abstract perspective in sections 12.2 and 12.3 respectively. In section 12.4, I conclude the thesis with a number of directions for future research.

12.2 Semi-lexicality and grammaticalisation

In this section I discuss the consequences that the theoretical proposal of this thesis has for our understanding of the early steps of grammaticalisation. The main proposal of this thesis is that there are two consecutive stages of semi-lexicality, which have different underlying syntactic structures. These are the very early stages of a grammaticalisation path, before a given vocabulary item has acquired a syntactic feature (and thus has become functional). This proposal leads me to adapt the grammaticalisation path from lexical to functional vocabulary item that was proposed by Hopper and Traugott (1993:120). It is represented in (1).

(1) $A_{lexical} > B_{functional}/A_{lexical}$ $(> B_{functional})$

The reasoning behind this grammaticalisation path is given in the following citation:

> There has in the past been a tendency to think of change in terms of 'A uniformly > B'. Given such an approach, divergence might be an unlikely characteristic. However, as we have noted, change must always been seen in terms of variation, and the formula for change should therefore be 'A > A/B > B'. Even so, it still needs to be stated that it is by no means inevitable that A will disappear. A and B may instead each go their own ways and continue to coexist as divergent reflexes of a historically single form over many centuries, even millennia. [...] The formula should therefore ideally be further modified to 'A > B/A (> B)'. (Hopper and Traugott 1993:120)

I adopt Hopper & Traugott's claim that the lexical vocabulary item does not necessarily have to disappear from the language when the functional vocabulary item has emerged. What I do want to change in their grammaticalisation path, however, is the number of stages from the lexical vocabulary item ($A_{lexical}$) to the functional one ($B_{functional}$). As I have claimed in this thesis, there are two stages in which a lexical vocabulary item is used semi-lexically, before it (potentially) transitions into a functional vocabulary item. I therefore adapt Hopper and Traugott (1993)'s grammaticalisation path to one with four rather than three stages. It is schematically presented in Table 12.1.

Stage	Vocabulary items
Stage 0	$A_{lexical}$
Stage I	$A_{lexical}$ + semi-lexical use_{stageI} of $A_{lexical}$
Stage II	$A_{lexical}$ + semi-lexical $use_{stageII}$ of $A_{lexical}$
Stage III	$B_{functional}$ (+ $A_{lexical}$)

Table 12.1: Grammaticalisation path (expanded version)

In stage 0, no grammaticalisation has taken place yet: there is only the lexical vocabulary item. In stage I, the lexical vocabulary item can be used lexically and semi-lexically. When it is used semi-lexically, the underlying syntactic structure is a root Merged very low in the functional domain of another root. In stage II, the lexical vocabulary item can also be used lexically and semi-lexically, but when it is used semi-lexically, the underlying syntactic structure is a root Merged in a separate workspace with a functional head. In stage III, the semi-lexical use of stage II has led the language learner to postulate the syntactic feature of that functional head on the vocabulary item itself. At this stage, a functional vocabulary item (i.e. the spell-out of a syntactic feature) has emerged and it is a separate vocabulary item from the lexical vocabulary item (which spells out a featureless root). As Hopper and Traugott (1993) note, the lexical vocabulary item can be lost at this stage, but it can just as easily be retained, hence the brackets around $\mathrm{A}_{lexical}$ in Stage III.

With respect to stage I of this expanded grammaticalisation path it is important to note that the semantic features of the lexical vocabulary item should be compatible with the interpretation of the syntactic feature it might eventually grammaticalise into. In both case studies, we have seen semi-lexical use of a specific class of verbs, namely motion and posture verbs. Even though case study I only focused on the semi-lexical use of the Dutch posture verb *zitten* 'sit', the other two posture verbs *staan* 'stand' and *liggen* 'lie', as well as the motion verb *lopen* 'walk' can also be used to indicate progressive aspect of a lexical verb. In case study II, we have seen that the Afrikaans motion verb *loop* 'walk' and the three posture verbs, *sit* 'sit', *staan* 'stand' and *lê* 'lie' can be used to indicate progressive/durative aspect of a lexical verb. In addition, *loop* can also be used to indicate andative aspect. The fact that these verbs are a target for grammaticalisation and as such are used to indicate a type of aspect, is not restricted to Dutch and Afrikaans. It is a development that can be found in many (related as well as unrelated) other languages as well (Kuteva 1999). This should not come as a surprise, given that the lexical semantics of all these verbs are highly compatible with a syntactic specification of progressive/durative aspect. That is, all posture verbs have a notion of duration in their event semantics, given that sitting, standing and lying are all static events Newman (2002). The motion verb *lopen*/*loop* has a notion of iteration and repetitive movement, which aligns well with progressive/durative aspect as well (Ross 2016). This verb furthermore also semantically contains the notion of movement going away from the deictic center/the speaker, which is very compatible with andative aspect (Ross 2016). All of this illustrates that the lexical semantics of a vocabulary item should be highly compatible with the syntactic feature it might grammaticalise into in order for the grammaticalisation process to take off. As Biberauer (2016:2) phrases it, the grammaticalisation of syntactic features 'piggy-backs' on semantic features (cf. also Zeijlstra 2008). Furthermore, the vocabulary item with the most general instantiation of the relevant semantics, i.e. with the least additional or specific semantic features, is the one that will become grammaticalised (Lehmann 1985:303–318, Traugott and Heine 1991:7–8, Heine et al. 1991:221–222, and Hopper and Traugott 1993:154–155). For example, in the case of motion verbs, verbs with the meaning GO or WALK are the ones that become grammaticalised, but not RUN, SPRINT, STROLL, et cetera. In the case of posture verbs, it is always one or more of the cardinal posture verbs that become grammaticalised, namely SIT, *stand*, and *lie*, but not more specific ones like CROUCH, or KNEEL.

Another important note with respect to Table (1) is the fact that multiple grammaticalisation paths can start from one and the same lexical vocabulary item. This is what we have seen in the second case study for the Afrikaans motion verb *loop* 'walk'. First, the grammaticalisation path started from the lexical vocabulary item *loop* in the direction of andative aspect, the semi-lexical use of *loop* which is now in the second stage of semi-lexicality (i.e. Stage II in Table (1)). This development started much earlier than that of the progressive use of *loop* (Roberge 1994). This means that for the progressive use of *loop*, the grammaticalisation process started from scratch again. This process is now in the first stage of semi-lexicality for most speakers, and in the second stage for a small group of speakers. These developments involve two unconnected grammaticalisation paths, with the one for the andative use of *loop* having started much earlier than the one for the progressive use of *loop*.

In subsection 5.4.6 of case study I, I have argued that the transition from the first to the second stage of semi-lexicality can go quite fast when two enforcing factors are present. The first enforcing factor concerns frequency. When the semi-lexical use is be frequent enough compared to the lexical use of the vocabulary item, this enforces further grammaticalisation. The second enforcing factor is robust existence in the PLD for the existence a given syntactic feature [F] in the language from functional vocabulary items bearing that feature. We have seen this for *hoeven* 'must', one of the restructuring verbs investigated in case study I. The semi-lexical use of this verb is much more frequent that the lexical use. When used semi-lexically, the interpretation of *hoeven* is that of a modal (of necessity). Given that there is a clear group of modal verbs in Dutch (with specifically the modal *moeten* 'must' indicating necessity), the language learner has robust evidence for the existence of a [MOD]-feature in the language. Thus, the combination of the high relative frequency of the semi-lexical use of *hoeven* and the evidence in favour of an independent [MOD]-feature in the language makes enforce *hoeven*'s grammaticalisation from the first to the second stage of semi-lexicality at a rapid pace. This is corroborated by Van de Velde (2017)'s finding that the semi-lexical use of *hoeven* has shown a dramatic decrease in selecting a *te*-complement over the last fifty years. As I have argued in subsection 5.4.3, there is no position in which *te* can be spelled out in the syntactic structure of the second stage of semi-lexicality. In the same subsection, I have also argued that when these two enforcing factors are absent, the grammaticalisation path halts in the first stage of semi-lexicality. I have argued that this is the case for the semi-lexical use of *zitten*. The lexical use of this verb has a much higher frequency than that of the semi-lexical use. Furthermore, there is no evidence for the language learner to postulate the existence of a [PROG]-feature in the language.

Summing up, in this thesis I have proposed a more refined grammaticalisation path than Hopper and Traugott (1993), and I have also discussed two factors that can influence the rate at which a given vocabulary item grammaticalises from one stage to the next. Proposing a formal analysis for these early stages of grammaticalisation is something that has not yet been done within the generative framework. In much of the generative work on grammaticalisation (see among others Roberts and Roussou 2003, Gelderen 2011, the main concern is how to formalise feature loss of vocabulary items, but with a focus on later stages of grammaticalisation processes. In contrast, my proposal is a step towards formally understanding the stages of grammaticalisation leading up to the adoption of a syntactic feature by a lexical item. As such, it fills an important gap in our understanding of early grammaticalisation.

12.3 Patterns of morphosyntactic optionality

Language variation and optionality have been extensively discussed at least since the sixties (see, among many others, Labov 1966, Weinreich et al. 1968, and more recently Labov 1994, Kroch 2001, Henry 2002). (Morpho)syntactic optionality is a topic that has recently received a more focused attention within the Minimalist Program as well, with publications such as Adger and Smith (2005, 2010), Adger (2006, 2014, 2016), Biberauer and Richards (2006), Richards (2008), Barbiers (2014), Nowenstein (2014), Tortora (2014), Tamminga et al. (2016) and Thoms et al. (2019). In this section, I discuss how some, but not all, of the patterns of morphosyntactic optionality the two case studies in this thesis have revealed are due to the semi-lexical use of the relevant vocabulary items. The aim of this section is to show that the patterns of optionality found in this thesis cannot be explained by one uniform theory of optionality in semi-lexical restructuring contexts in Dutch and Afrikaans. In contrast, the optionality patterns can only be analysed by looking into the precise underlying syntactic structures of the different configurations presented the two case studies. In so doing, we also need to take into account language specific factors that have influenced or caused part of the optionality.

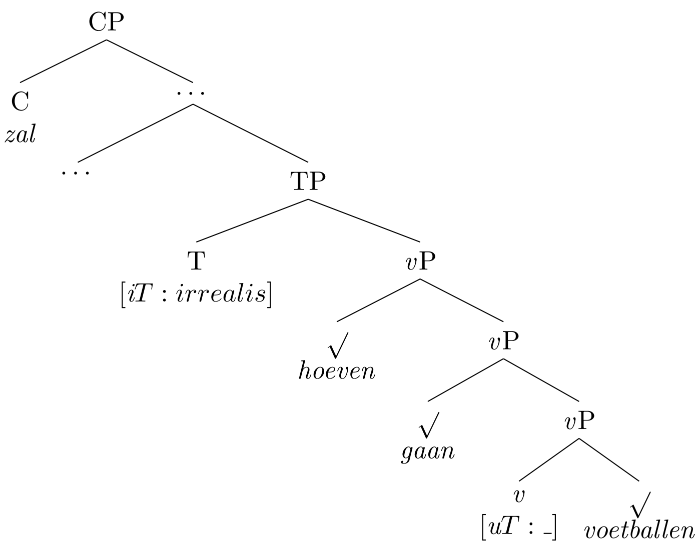
The optionality patterns discussed in both case studies are not all caused by the same factors. Six such factors can be distinguished, namely:

1. Ongoing grammaticalisation (Dutch *hoeven* 'need', Afrikaans PC verbs $loop_{progressive}$ 'walk' and *staan* 'stand')
2. A syntactic structure with multiple positions for a feature to be spelled out in (Dutch *hoeven* 'need')
3. An optional alternative process of Agree and valuation in syntax (all Dutch non-finite verb clusters)
4. The availability of two different syntactic structures, due to language contact and standardisation (Afrikaans V2 PC constructions)
5. Overgeneralisation of the morphosyntax of one of the semi-lexical uses of a vocabulary item to the morphosyntax of the other semi-lexical use of that vocabulary item (Afrikaans PC verb $loop_{andative}$)
6. A more general optionality of a morphosyntactic element in the language (Afrikaans *ge-* in embedded PC constructions)

I will briefly recapitulate how these different factors cause morphosyntactic optionality. The first is ongoing grammaticalisation. We have seen this as a factor responsible for morphosyntactic optionality in case study I in the cluster type with *hoeven* (cluster type II), and in case study II in PC constructions with $loop_{progressive}$ and *staan*. The main gist of the analysis in both case studies is that for a group of speakers, these verbs are still in the first stage of semi-lexicality, whereas for others they are already in the second stage. The different underlying syntactic structures of the two stages result in a different morphosyntax. Thus, this factor is directly connected to the semi-lexical use of the verbs involved.

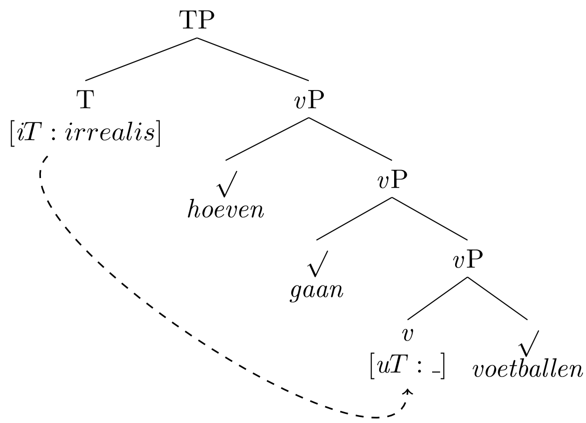
The second factor is a configuration where there are multiple positions for a feature to be spelled out in. We have seen this in the first case study, for the underlying structure of the cluster type in which *hoeven* is used (cluster type II) when it is in the first stage of semi-lexicality. This syntactic structure is repeated here in (2).

(2) **Cluster type II: structure (semi-lexical stage I)**



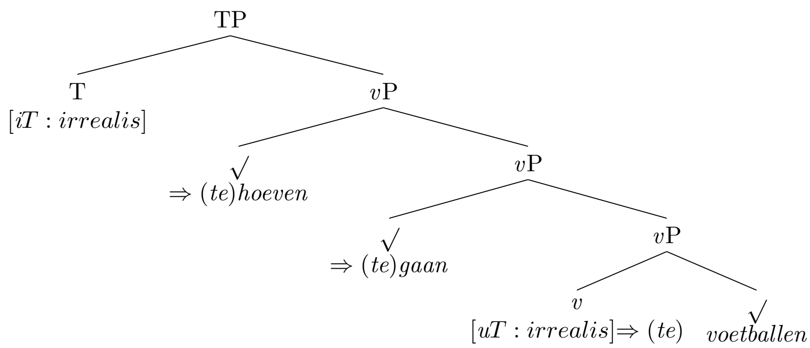
Recall from section 5.3.3 of the first case study that I assume Upward Agree, and propose that *te* can be the spell out of a *u*T-feature that is valued for *irrealis*. The feature valuation of cluster type II in the first stage of semi-lexicality is repeated in (3).

(3) **Cluster type II: valuation (semi-lexical stage I)**



The unvalued uT-feature on v probes up for a iT-feature that can give it a value. Given that *hoeven* and *gaan* 'go' are used semi-lexically, they are featureless roots. As a result, they cannot value the uT-feature on v. This valuation is done by [iT:irrealis] on T, and it results in a spell-out conflict: as v projects three times in the structure, there are three possible positions *te* can be spelled out in, as illustrated again in (4).

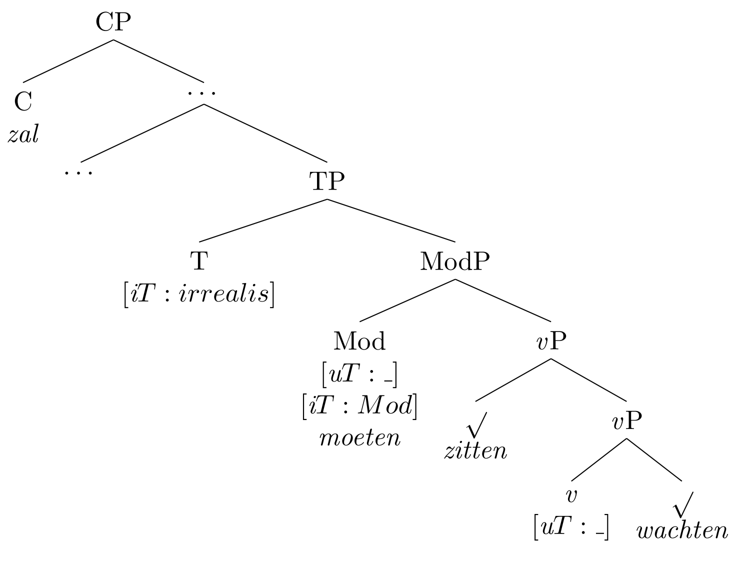
(4) **Cluster type II: spell out (semi-lexical stage I)**



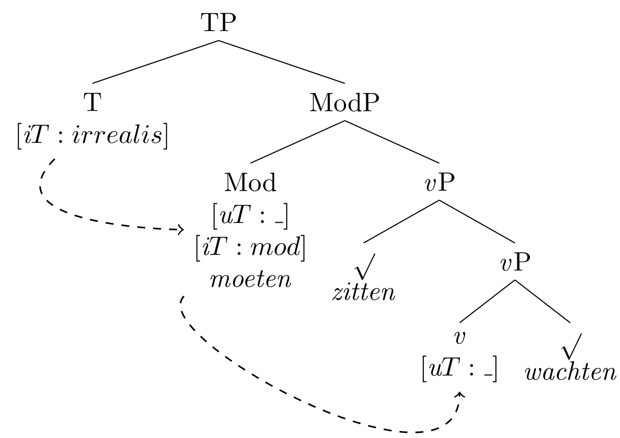
Given that there are three possible positions for *te* to be spelled out in, there is a high degree of morphosyntactic optionality in the position of *te* in this cluster type. In addition, I have argued that another solution to this spell-out conflict is to not spell out *te* at all, resulting in *te*-drop in this cluster type. The multiple spell out positions in this cluster type are a side-effect of the semi-lexical restructuring configuration of this cluster, and as such also directly related to semi-lexicality.

The third factor causing morphosyntactic optionality is an alternative Agree and valuation process in syntax. This factor is not directly related to semi-lexicality, as it is found in all Dutch non-finite cluster types, including cluster types Ia and Ib in which the *te*-selecting verb is not a semi-lexical. I repeat one illustration of this factor here, namely in the cluster type in which *zitten* is used (cluster type III). The syntactic structure for this cluster type is repeated in (5). The standard process of Agree and feature valuation in this cluster type is illustrated in (5), in which the [*u*T]-feature on *v* Agrees with the [*i*T:Mod]-feature on Mod. However, as I have argued in subsection 5.4.1, there is also a more marked, alternative process of Agree and valuation possible. In this alternative process the unvalued [*u*T]-feature on *v* Agrees with the unvalued [*u*T]-feature on Mod rather than the [*i*T:Mod]-feature on that head. This creates a link between the [*u*T]-feature on *v* and the one on Mod, as illustrated in (5).

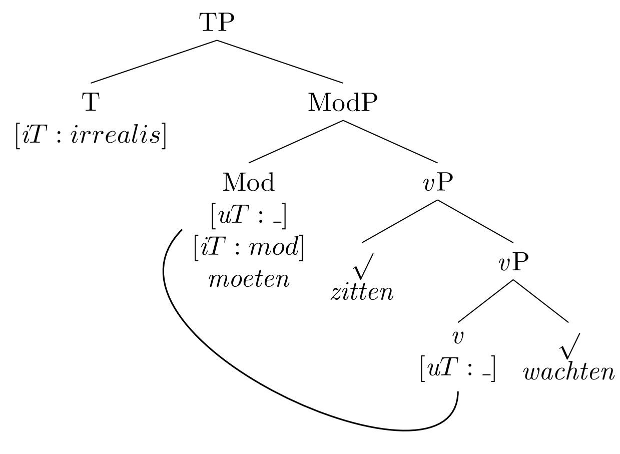
(5) **Cluster type III: structure (semi-lexical stage I)**



(6) **Cluster type III: valuation (semi-lexical stage I)**

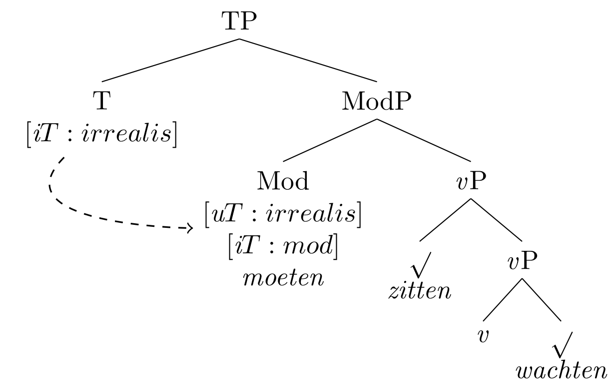


(7) **Cluster type III: alternative valuation step one**

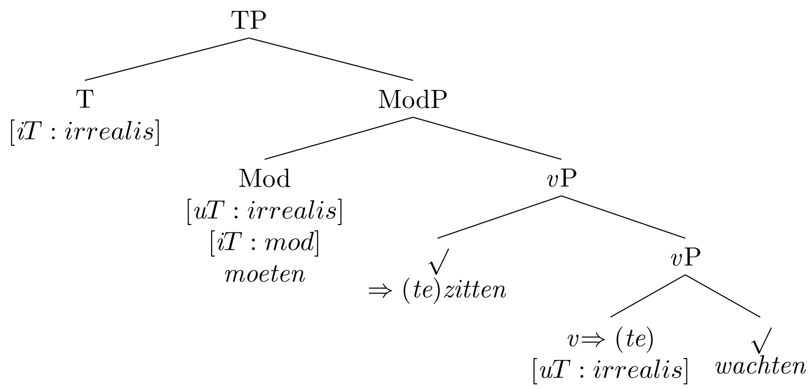


When this [*u*T]-feature on Mod Agrees with the higher [*i*T:*irrealis*]-feature on T, both the [*u*T]-feature on Mod and the [*u*T]-feature on *v* get valued for *irrealis*, as illustrated in (9). Given that an [*u*T:irrealis]-feature on *v* can be spelled out as *te*, and given that *v* projects twice in this cluster, the alternative Agree and valuation process results in two possible positions for *te* to be spelled out in, shown in (8).[1]

(8) **Cluster type III: alternative valuation step two**



(9) **Cluster type III: alternative spell out**



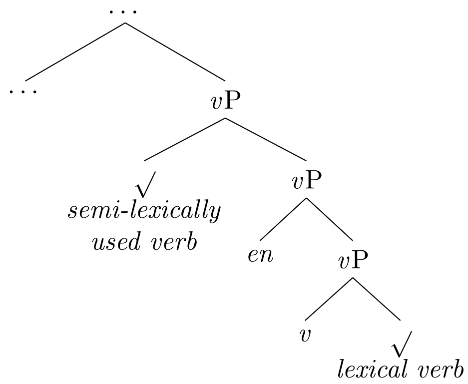
Thus, an alternative process of Agree and valuation can lead to morphosyntactic optionality.

The fourth factor causing optionality is the availability of two different syntactic structures, due to language contact and standardisation. We have seen this with the V2 PC configurations in the Afrikaans case study. This factor is related to semi-lexicality, because only in the first stage of semi-lexicality are two different syntactic structures available. However, the morphosyntactic optionality which results from these two alternative structures is not caused by the semi-lexicality of the involved verbs per se, but rather by a complex diachronic development of the Afrikaans PC constructions, in which language

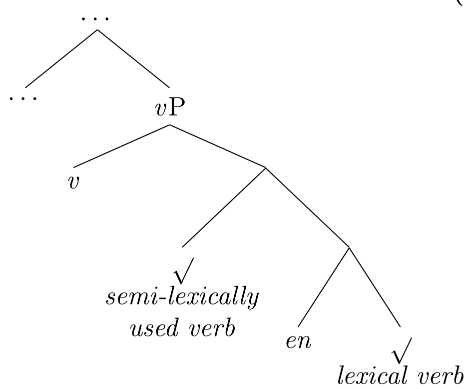
[1] Recall that the [uT]-feature on Mod in (8) cannot be spelled out as *te*, because this is not a v-head.

contact and standardisation have played a big role. That is, in subsection 10.7 of the Afrikaans case study, I have proposed that this complex diachronic development has resulted in the availability of both a 'traditional' and an 'innovative' syntactic structure for PC constructions in the first stage of semi-lexicality. These two structures are repeated here in (10) and (11).

(10) **The 'traditional' PC construction (semi-lexical stage I)**



(11) **The 'innovative' PC construction (semi-lexical stage I)**



In V2 PC constructions, the 'traditional' construction results in normal V2, an example of which is given in (12).

(12) Hoekom **loop** Jan heeldag piesangs **en** **eet**?
why walk.FIN Jan all.day bananas and eat
'Why is Jan (walking and) eating bananas all day long?'

In the 'innovative' construction, we find so-called 'quirky' V2, an example of which is given in (13).

(13) Hoekom **loop** **en** **eet** Jan heeldag piesangs?
why walk.FIN and eat Jan all.day bananas
'Why is Jan (walking and) eating bananas all day long?'

The availability of two different syntactic structures can thus also cause morphosyntactic optionality.

The fifth factor causing optionality is the overgeneralisation of the morphosyntax of one of the semi-lexical uses of a vocabulary item to the morphosyntax of the other semi-lexical use of that vocabulary item. We have seen this in case study II for the Afrikaans PC verb $loop_{andative}$. As discussed in section 10.6 of that case study, the diachronic development of this verb is separate from that of the other PC verbs. As a result of standardisation processes, the morphosyntactic structure of the other PC verbs has been changed from a structure without the linker *en* to one that includes this linker. In subsection 10.8.1, I have argued that some speakers have overgeneralised the presence of *en* to the morphosyntactic structure of PC constructions with $loop_{andative}$. The optionality of *en* in these PC constructions is thus not related to the semi-lexical use of this verb, but to a process of overgeneralisation of the morphosyntactic structure of other PC verbs to the one containing $loop_{andative}$.

The sixth and final factor responsible for optionality is a more general optionality of a morphosyntactic element in the language. We have seen this in case study II for the Afrikaans verbal marker *ge-*. In subsection 10.5.3 of that case study, I have argued that *ge-* is an optional element in Afrikaans PC constructions embedded under a perfective auxiliary. This is due to a phonological rule which makes *ge-* optional in unstressed-stressed patterns, and thus also in PC constructions, for example *loop kóóp*. A potential additional factor in the optionality of *ge-* is the fact that the Afrikaans temporal auxiliary *het* 'have', which is present in embedded PC constructions, has developed into a marker of past tense. Given that *ge-* has undergone a similar development (as argued in subsection 10.5.1), the optionality of *ge-* is enforced by the presence of *het* in the PC construction as well.

Summing up, different analyses were needed to account for different types of morphosyntactic optionality patterns throughout the two case studies. This thesis thus shows that even within a single syntactic domain (i.e. the verbal domain) in two closely related languages (i.e. Dutch and Afrikaans), and furthermore in very specific contexts (i.e. semi-lexical restructuring), different factors play a role in patterns of morphosyntactic optionality, making a uniform morphosyntactic account of such optionality problematic. This furthermore shows that even though in principle a uniform analysis is worth striving for, when one deals with very complex data sets, a combination of various detailed analyses is needed to do justice to the variation and optionality found in those data.

12.4 Conclusions and future prospects

The main theoretical contribution of this thesis has been the formalisation of the early stages of grammaticalisation (semi-lexical stage I and II) in restructuring contexts in Dutch and Afrikaans. This proposal has been illustrated and elaborated on in two case studies, one dealing with semi-lexicality in non-finite verb clusters in Dutch (case study I), and one with semi-lexicality in pseudo-coordination constructions in Afrikaans (case study II). In both case studies, a large amount of new data on morphosyntactic variation and optionality in the two languages was presented. This morphosyntactic variation and optionality was statistically analysed with both descriptive and exploratory statistical techniques, which laid the foundation for specific and detailed formal analyses for both languages.

In this final section, I close off the thesis with a number of directions for future research. I first briefly recapitulate the future prospects presented in the conclusion and outlook chapters of the two case studies (Chapters 6 and 11 respectively). After that, I propose two broader directions for future research related to the topic of this thesis.

In Chapter 6 of the first case study, I have discussed three future prospects related to the presence and placement of *te* and semi-lexical restructuring. First, I suggested that there are three Dutch restructuring verbs that seem to be developing from lexical to semi-lexical. These verbs are *durven* 'dare', *beginnen* 'begin' and *proberen* 'try'. Given that semi-lexical restructuring configurations can result in optionality in the presence and placement of *te*, I proposed that a future questionnaire study focused on non-finite verbs clusters with these three verbs could be used as a testing ground for the analysis presented in that case study. Second, I discussed the fact that *te* is obligatory when the two semi-lexically used verbs of the case study, *hoeven* 'need' and *zitten* 'sit' are the finite verb in a Verb Second configuration in root clauses, whereas it is optional or even preferably absent in when these verbs are non-finite and part of a verb cluster at the end of the clause. I hope that future research can result in a better understanding of this finite/non-finite distinction, which could also potentially contribute to a better understanding of the Verb Second phenomenon in West-Germanic languages. Third, I discussed the fact that the interaction between the presence and placement of *te* and different word order patterns in non-finite verb clusters would be a fruitful future direction of research as well, and I furthermore pointed out that I have already collected data of this nature.

In Chapter 11 of the second case study, I have presented one main direction for future research, namely an in-depth investigation into the complex diachronic development of the Afrikaans PC constructions, in which sociolinguistic factors should be considered as well. The diachronic development and present-day morphosyntax of these PC constructions has been strongly influenced by language contact and standardisation. Given that my formal analysis of this case study builds on these facts, it would be very fruitful to investigate the change of the morphosyntax of PC constructions in Afrikaans prose of the previous century, as well as the type of instruction children get in schools nowadays regarding the morphosyntax of PC constructions.

A broader research question emerging from this thesis is whether the presence of restructuring in a given language makes it easier for lexical verbs to grammaticalise and be used semi-lexically compared to languages in which restructuring does not exist. My hypothesis is that this is indeed the case. Consider in this respect the Scandinavian languages, in which pseudocoordination with motion and posture verbs is attested as well (see among others Lødrup 2002, Wiklund 2007, Biberauer and Vikner 2017). An example from Danish is given in (14).

(14) Han **sidder og smiler**.
he sits and smiles
'He is sitting and smiling.' (Biberauer and Vikner 2017:80)

In these pseudocoordination constructions, however, the lexical semantics of the motion or posture verb is completely retained, and interestingly, these languages do not have restructuring. I believe the development of pseudocoordination constructions in which the motion or posture verb becomes more grammaticalised might be facilitated by the presence of restructuring in the language. Given the absence of restructuring in Scandinatian, it would be interesting to investigate the stable non-grammaticalised state of the pseudoocordination constructions in these languages. This could be done by investigating the degree of semantic bleaching of the motion/posture verb as well as their morphosyntax. In addition to this, a comparison of the Dutch and Afrikaans restructuring constructions with motion and posture verbs with motion verb constructions in Italian varieties would also be worthwhile (cf. Cardinaletti and Giusti (2001)). Comparing the Italian motion verb constructions with the Afrikaans and Dutch ones would be very interesting because Italian also has restructuring (Rizzi 1978, Cinque 2001). Thus, the impact of restructuring on the development of constructions with motion and posture verbs could also be investigated from a combined Germanic/Romance perspective.

A second broader direction for future research is the investigation of differences between the verbal domains of Dutch and Afrikaans from a more general point of view. In this thesis, the two case studies have shown that even though both languages show semi-lexical restructuring with motion and posture verbs, the underlying structures of these types of restructuring configurations is not exactly the same in the two languages. Furthermore, while the Dutch posture verb *zitten* 'sit' is in the first stage of semi-lexicality, some of the Afrikaans PC verbs are in the second stage for a group of speakers (i.e. $loop_{progressive}$ 'walk' and *staan* 'stand'). In addition, Dutch never developed an andative use of *lopen* 'walk', while Afrikaans did. However, the verbal domains of Dutch and Afrikaans not only differ with respect to the semi-lexical use of motion and posture verbs. Let me illustrate this with two ways in which Afrikaans modals differ from Dutch modals. First, the Afrikaans modal *moet* 'must' is used in negative imperatives in the contracted form *moe-nie* 'must not', as illustrated in (15) (Biberauer 2018).

(15) **Moe-nie** die deur oopmaak nie!
must-not the door open.make POL
'Don't open the door!' (Biberauer 2018:12)

The construction with *moenie* 'must not' is in fact the only possible form for the negative imperative in Afrikaans. In Dutch, the modal *moeten* cannot be used to form a negative imperative (16-a). It can only be used to indicate declarative command, as illustrated in (16-b).

(16) a. ***Moet** de deur niet openmaken!
must the door not open.make
b. ***Je*** **moet** de deur niet openmaken!
you must the door not open.make
'Don't open the door!'

Instead, Dutch has two negative imperative constructions that do not make use of the modal *moeten*. They are illustrated in (17-a) and (17-b).

(17) a. **Maak** de deur niet open!
make the door not open
'Don't open the door!'
b. De deur niet **openmaken**!
the door not open.make
'Don't open the door!'

Neither of these negative imperatives is possible in Afrikaans (18).

(18) a. ***Maak** die deur nie oop nie!
make the door not open POL
b. *Die deur nie **oopmaak** nie!
the door not open.make POL
(Theresa Biberauer, p.c.)

Thus, while in Afrikaans one of the modals, *moet*, is used extensively in negative imperative constructions, and is in fact the only way to form a negative imperative, in Dutch negative imperatives the modal *moeten* is not used.

A second difference between the Afrikaans and Dutch modals is the fact that Afrikaans modals are used in irrealis and subject-oriented contexts, whereas in Dutch this is ungrammatical Biberauer et al. (2019). Two examples from Afrikaans are given in (19) and (20), one with the modal *moet* 'must' and one with *kan* 'can'.

(19) Hy wag dat Ludo **moet** opstaan.
he wait that Ludo must up.stand
'He waits for Ludo to get up.'

(20) Dit is werklik 'n risiko dat die myn **kan** verdwyn.
it is really a risk that the mine can disappear
'There is really a risk that the mine could dissapear.' (Biberauer et al. 2019:13–14)

The Dutch versions are given in (21) and (22). The use of a modal here is ungrammatical.

(21) Hij wacht erop dat Ludo opstaat / ***moet** opstaan.
He waits it.for that Ludo up.stand / must up.stand
'He waits for Ludo to get up.'

(22) Het is werkelijk een risico that de mijn verdwijnt / ***kan** verdwijnen.
it is really a risk that the mine disappears / can disappear
'There is really a risk that the mine could dissapear.'
(Biberauer et al. 2019:13–14)

It would be a fruitful direction for future research to do a broader in-dept comparison between the Dutch and Afrikaans verbal domain, focusing on the exact differences between the modal verbs, the auxiliary verbs (cf. section 10.4 of case study II on the Afrikaans auxiliary *het* 'have'), and the semi-lexically used verbs. This could lead to a deeper understanding of which aspects of the verbal domain in Afrikaans are still similar to those of Dutch, and which ones have developed in new directions compared to Dutch in particular, and the verbal domain of European West-Germanic languages in general.

Appendix A

A1: Instruction text of Dutch questionnaire study I (in Dutch)

Introductie
Beste meneer/mevrouw,

Hartelijk dank dat u wilt deelnemen aan dit onderzoek binnen het project Quality and Quantity in Linguistics, dat uitgevoerd wordt aan KU Leuven. Wij wijzen u erop dat deelname geheel anoniem is, en dat de resultaten enkel binnen het kader van dit onderzoek gebruikt zullen worden. Mocht u een vraag hebben met betrekking tot dit onderzoek, kunt u een e-mail sturen naar [email address].

Het onderzoek bestaat uit het invullen van een enquête. Voordat we daarmee beginnen, willen we u eerst wat instructies geven over hoe dat in zijn werk gaat. Klik op 'volgende' om naar de instructies te gaan.

Instructie
Moedertaalsprekers hebben altijd een erg duidelijk 'gevoel' over wat een mogelijke of onmogelijke zin is in hun moedertaal, zelfs als ze daar nooit een specifieke regel voor hebben geleerd. Dat geldt niet alleen voor de uitspraak van een zin, of de gepastheid van een zin in een bepaalde situatie, maar ook voor de structuur van een zin. Bijvoorbeeld, als moedertaalspreker van het Nederlands voelt u aan dat de structuur van de volgende zin mogelijk is in het Nederlands:

(1) Ik denk dat mijn zus een boek koopt.

Het gaat daarbij niet over of u wel of geen zus heeft, maar alleen over dat u voelt dat de structuur van de zin in orde is. Daarentegen, bij een zin als de volgende weet u intuïtief dat de structuur van de zin géén mogelijke structuur is in het Nederlands:

(2) Ik denk dat mijn zus koopt een boek.

Voor dit onderzoek willen we u vragen een aantal zinnen op dezelfde manier te bekijken. U krijgt straks 65 zinnen één voor één te zien. We willen u vragen om voor iedere zin aan te geven of de zin een mogelijke of onmogelijke zin is in het Nederlands zoals dat in uw directe omgeving (familie, vrienden, dorp of stad) wordt gesproken. Concentreer u op de structuur van de zin, en negeer gevoelens over hoe groot of klein de kans is dat u die zin ook echt in een gesprek gebruikt.

Voor sommige zinnen zal uw gevoel over de structuur minder duidelijk zijn. U kunt bijvoorbeeld het gevoel hebben dat een zin op zich wel mogelijk is, maar dat de structuur wat onnatuurlijk of ongemakkelijk is. We gaan daarom werken met een schaal van 1-5. Bij iedere zin stellen we u de vraag:

Is dit een mogelijke zin in het Nederlands zoals dat in uw directe omgeving wordt gesproken?

Als u op de vraag 'zeker' antwoordt, geeft u de zin een 5. Als u op de vraag 'zeker niet' antwoordt, geeft u de zin een 1. U kunt ook een 4, 3 of 2 toekennen. Als u echt geen gevoel heeft over de mogelijkheid van de zin, klikt u op 'weet ik niet'. Probeer die optie zo weinig mogelijk te gebruiken. Als u uw oordeel wilt toelichten of andere opmerkingen wil geven over de zin, kunt u dat doen in het lege veld onder de zin.

Lees iedere zin hardop voor en klik direct daarna uw oordeel (1-5 of 'weet ik niet') aan. Uw eerste reactie is het belangrijkst. Er zijn geen goede of slechte antwoorden; het enige wat van belang is, is uw gevoel over de mogelijkheid van de zin.

Het invullen van de enquête duurt ongeveer 20 minuten. Soms lijken de zinnen erg op elkaar. Vergeet dus niet iedere zin weer hardop te zeggen voordat u de zin beoordeelt. Het is mogelijk niet de meest interessante enquête om in te vullen voor u als deelnemer, maar uw antwoorden zijn van groot belang voor dit onderzoek. We hopen daarom ook dat u bereid bent de enquête helemaal af te maken.

We zijn erg dankbaar dat u wilt deelnemen aan dit onderzoek!

A2: List of practice items, fillers, test items and background questions of Dutch questionnaire study I (in Dutch)

Practice items

1. Sylvia vertelt dat haar vriend een nieuwe auto gaat kopen. (*condition*: grammatical)
2. Geert zegt dit jaar nog niets te hebben gehoord van zijn vrienden uit Frankrijk. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* on correct position)
3. Sarah gisteren heeft al haar vrienden een klein cadeautje gegeven. (*condition*: ungrammatical)
4. Saskia zegt afgelopen weekend weer veel in haar tuin hebben kunnen te werken. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* displaced)
5. Kasper hoopt vanaf deze maand vaker naar de volleybaltraining kunnen te komen. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* displaced)

Fillers

1. Jan hoopt volgend jaar met zijn nieuwe baan meer geld te verdienen. (*condition*: grammatical)
2. Jan hoopt volgend jaar met zijn nieuwe baan te verdienen meer geld. (*condition*: ungrammatical)
3. Sarah zegt dat haar zus dat boek heeft willen kopen. (*condition*: grammatical)
4. Sarah zegt dat haar zus dat boek willen heeft kopen. (*condition*: ungrammatical)
5. Thomas zegt dat zijn collega niet heeft willen luisteren naar zijn advies. (*condition*: grammatical)
6. Thomas zegt dat zijn collega niet heeft gewild luisteren naar zijn advies. (*condition*: ungrammatical)
7. Sylvia vertelt dat haar vriend een nieuwe auto gaan kopen. (*condition*: ungrammatical)

8. Robert belooft zijn kinderen om vandaag vroeg naar huis te komen. (*condition*: grammatical)
9. Robert belooft zijn kinderen om vandaag vroeg naar huis komt. (*condition*: ungrammatical)
10. Jan zegt dat hij dit weekend zijn moeder wil bezoeken. (*condition*: grammatical)
11. Jan zeggen dat hij dit weekend zijn moeder wil bezoeken. (*condition*: ungrammatical)
12. Sarah heeft gisteren al haar vrienden een klein cadeautje gegeven. (*condition*: grammatical)
13. Thomas heeft de afgelopen maand erg veel nieuwe kleren gekocht. (*condition*: grammatical)
14. Thomas heeft de afgelopen maand erg veel nieuwe kleren kopen. (*condition*: ungrammatical)
15. Steven moet wel heel goed verdienen om die auto te hebben kunnen kopen. (*condition*: control sentence with *te* selected by *om*; *te* in correct position)
16. Steven moet wel heel goed verdienen om die auto hebben te kunnen kopen. (*condition*: control sentence with *te* selected by *om*; *te* lowered to V2)
17. Steven moet wel heel goed verdienen om die auto hebben kunnen te kopen. (*condition*: control sentence with *te* selected by *om*; *te* lowered to V3)
18. Steven moet wel heel goed verdienen om die auto hebben kunnen kopen. (*condition*: control sentence with *te* selected by *om*; *te* absent)
19. Steven moet wel heel goed verdienen om die auto te hebben te kunnen kopen. (*condition*: control sentence with *te* selected by *om*; *te* doubled)
20. Steven moet wel heel goed verdienen om die auto te hebben kunnen te kopen. (*condition*: control sentence with *te* selected by *om*; *te* doubled)
21. Steven moet wel heel goed verdienen om die auto hebben te kunnen te kopen. (*condition*: control sentence with *te* selected by *om*; *te* doubled)
22. Erik vind het vervelend dat zijn moeder de hele dag loopt te klagen. (*condition*: additional filler)
23. Erik hoopt dat zijn moeder niet weer gaat lopen te klagen bij de buren. (*condition*: additional filler)

24. Erik hoopt dat zijn moeder niet weer gaat lopen klagen bij de buren. (*condition*: additional filler)

25. Kasper vreest dat zijn broer op het feest weer gaat staan te pronken met zijn nieuwe vriendin. (*condition*: additional filler)

26. Kasper vreest dat zijn broer op het feest weer gaat staan pronken met zijn nieuwe vriendin. (*condition*: additional filler)

27. Saskia gaat ervan uit dat haar dochter op school weer gaat zitten te slapen. (*condition*: additional filler)

28. Saskia gaat ervan uit dat haar dochter op school weer gaat zitten slapen. (*condition*: additional filler)

29. Sandra is bang dat haar zus tijdens het etentje weer gaat liggen te zeuren over haar geldproblemen. (*condition*: additional filler)

30. Sandra is bang dat haar zus tijdens het etentje weer gaat liggen zeuren over haar geldproblemen. (*condition*: additional filler)

31. Geert belooft zijn studenten dat ze niet naar de les hoeven te komen als het sneeuwt. (*condition*: control sentence to check whether participants allow the use *hoeven* 'need')

32. Geert zegt aan zijn collega dat hij deze keer niet naar de vergadering hoeft te komen. (*condition*: control sentence to check whether participants allow the use *hoeven* 'need')

Test items

1. Anne zegt op haar comfortabele stoel te willen blijven zitten. (*condition*: cluster type Ia; *te* in correct position)

2. Anne zegt op haar comfortabele stoel willen te blijven zitten. (*condition*: cluster type Ia; *te* lowered to V2)

3. Anne zegt op haar comfortabele stoel willen blijven te zitten. (*condition*: cluster type Ia; *te* lowered to V3)

4. Anne zegt op haar comfortabele stoel willen blijven zitten. (*condition*: cluster type Ia; *te* absent)

5. Anne zegt op haar comfortabele stoel te willen te blijven zitten. (*condition*: cluster type Ia; *te* doubled, V1 and V2)

6. Anne zegt op haar comfortabele stoel te willen blijven te zitten. (*condition*: cluster type Ia; *te* doubled, V1 and V3)

7. Anne zegt op haar comfortabele stoel willen te blijven te zitten. (*condition*: cluster type Ia; *te* doubled, V2 and V3)
8. Stijn beweert met het geld van zijn erfenis die grote villa te hebben kunnen kopen. (*condition*: cluster type Ib; *te* in correct position)
9. Stijn beweert met het geld van zijn erfenis die grote villa hebben te kunnen kopen. (*condition*: cluster type Ib; *te* lowered to V2)
10. Stijn beweert met het geld van zijn erfenis die grote villa hebben kunnen te kopen. (*condition*: cluster type Ib; *te* lowered to V3)
11. Stijn beweert met het geld van zijn erfenis die grote villa hebben kunnen kopen. (*condition*: cluster type Ib; *te* absent)
12. Stijn beweert met het geld van zijn erfenis die grote villa te hebben te kunnen kopen. (*condition*: cluster type Ib; *te* doubled, V1 and V2)
13. Stijn beweert met het geld van zijn erfenis die grote villa te hebben kunnen te kopen. (*condition*: cluster type Ib; *te* doubled, V1 and V3)
14. Stijn beweert met het geld van zijn erfenis die grote villa hebben te kunnen te kopen. (*condition*: cluster type Ib; *te* doubled, V2 and V3)
15. Koen zal vanwege de winterstop vandaag niet te hoeven gaan voetballen. (*condition*: cluster type II; *te* raised to V1)
16. Koen zal vanwege de winterstop vandaag niet hoeven te gaan voetballen. (*condition*: cluster type II; *te* in correct position)
17. Koen zal vanwege de winterstop vandaag niet hoeven gaan te voetballen. (*condition*: cluster type II; *te* lowered to V3)
18. Koen zal vanwege de winterstop vandaag niet hoeven gaan voetballen. (*condition*: cluster type II; *te* absent)
19. Koen zal vanwege de winterstop vandaag niet te hoeven te gaan voetballen. (*condition*: cluster type II; *te* doubled, V1 and V2)
20. Koen zal vanwege de winterstop vandaag niet te hoeven gaan te voetballen. (*condition*: cluster type II; *te* doubled, V1 and V3)
21. Koen zal vanwege de winterstop vandaag niet hoeven te gaan te voetballen. (*condition*: cluster type II; *te* doubled, V2 and V3)
22. Door de nieuwe dienstregeling zal Pieter binnenkort nog langer op de trein te moeten zitten wachten. (*condition*: *te* raised to V1)
23. Door de nieuwe dienstregeling zal Pieter binnenkort nog langer op de trein moeten te zitten wachten. (*condition*: *te* raised to V2)

24. Door de nieuwe dienstregeling zal Pieter binnenkort nog langer op de trein moeten zitten te wachten. (*condition*: *te* in correct position)
25. Door de nieuwe dienstregeling zal Pieter binnenkort nog langer op de trein moeten zitten wachten. (*condition*: *te* absent)
26. Door de nieuwe dienstregeling zal Pieter binnenkort nog langer op de trein te moeten te zitten wachten. (*condition*: *te* doubled, V1 and V2)
27. Door de nieuwe dienstregeling zal Pieter binnenkort nog langer op de trein te moeten zitten te wachten. (*condition*: *te* doubled, V1 and V3)
28. Door de nieuwe dienstregeling zal Pieter binnenkort nog langer op de trein te moeten zitten wachten. (*condition*: *te* doubled, V2 and V3)

Background questions (in Dutch)

Hartelijk dank voor het invullen van de enquête!

Tot slot willen we u nog vragen enkele persoonlijke gegevens in te vullen. Deze gegevens kunnen ons helpen bij het interpreteren van de ingevulde enquêtes. Wij willen nogmaals benadrukken dat deze gegevens anoniem zijn, en dat ze enkel binnen het kader van dit onderzoek gebruikt zullen worden.

1. Wat is/zijn uw moederta(a)l(en)?
2. Spreekt u een dialect? Zo ja welk:
3. Wat is uw geslacht? (mannelijk, vrouwelijk, anders)
4. Wat is uw leeftijd?
5. Wat is uw hoogst behaalde opleidingsniveau? (primair, secundair, MBO, hbo/hogeschool, universiteit)
6. In welk land bent u opgegroeid? (Nederland, België, anders)
7. In welke plaats bent u opgegroeid?
8. Tot welke leeftijd heeft u daar gewoond?
9. Heeft u ooit voor een tijd in het buitenland gewoond? Zo ja, hoelang: en waar:
10. Wat is uw huidige woonplaats? Sinds wanneer woont u daar?
11. Werkt u verder dan 10 kilometer van uw woonplaats af? Zo ja, hoeveel kilometer?

Appendix B

B1: Instruction text Dutch questionnaire study II (in Dutch)

Beste deelnemer,

Voor dit onderzoek willen we u vragen een aantal zinnen te beoordelen. U krijgt straks 159 zinnen één voor één te zien.

We willen u vragen om voor iedere zin aan te geven of de zin een mogelijke of onmogelijke zin is in het Nederlands zoals dat in uw directe omgeving (familie, vrienden, dorp of stad) wordt gesproken. Concentreer u op de structuur van de zin, en negeer gevoelens over hoe groot of klein de kans is dat u die zin ook echt in een gesprek zou gebruiken.

Voor sommige zinnen zal uw gevoel over de structuur minder duidelijk zijn. U kunt bijvoorbeeld het gevoel hebben dat een zin op zich wel mogelijk is, maar dat de structuur wat onnatuurlijk of ongemakkelijk is. We gaan daarom werken met een schaal van 1 t/m 5. Bij iedere zin stellen we u de vraag:

Is dit een mogelijke zin in het Nederlands zoals dat in uw directe omgeving wordt gesproken?

Als u op deze vraag ‘zeker’ antwoordt, geeft u de zin een 5. Als u op de vraag ‘zeker niet’ antwoordt, geeft u de zin een 1. U kunt ook een 4, 3 of 2 toekennen. Als u echt geen gevoel heeft over de mogelijkheid van de zin, klikt u op ‘weet ik niet’. Probeer die optie echter zo weinig mogelijk te gebruiken. Als u uw oordeel wilt toelichten, kunt u dat doen in het lege veld onder de zin.

Lees iedere zin hardop voor en klik direct daarna uw oordeel (1-5 of ‘weet ik niet’) aan. Uw eerste reactie is het belangrijkst. Er zijn geen goede of slechte antwoorden; het enige wat van belang is, is uw gevoel over de mogelijkheid van de zin.

Als u wilt, kunt u de vragenlijst – die vrij lang is – in een drietal sessies afronden. Tijdens het invullen van de vragenlijst wordt dat duidelijk gemaakt.

Iedere sessie duurt ongeveer 20 minuten. Uw antwoorden worden automatisch opgeslagen. Soms lijken de zinnen erg op elkaar. Vergeet dus niet iedere zin weer hardop te lezen voordat u de zin beoordeelt. Het is mogelijk niet de meest interessante vragenlijst om in te vullen voor u als deelnemer, maar uw antwoorden zijn van groot belang voor dit onderzoek. We hopen daarom dan ook dat u bereid bent de vragenlijst helemaal af te maken.

We zijn erg dankbaar dat u wilt deelnemen aan dit onderzoek!

Mocht u inhoudelijke vragen hebben over het onderzoek of vragen over het Meertens Panel kunt u terecht bij de Afdeling communicatie.

Deze vragenlijst is op een groot aantal verschillende browsers op verschillende platforms getest. We kunnen helaas geen enkele technische ondersteuning bieden.

De vragenlijst begint als u op START klikt.

B2: List of practice items, fillers, test items and background questions of Dutch questionnaire study II (in Dutch)

Practice items

Practice round I

1. Sylvia vertelt dat haar vriend een nieuwe auto gaat kopen. (*condition*: grammatical)
2. Sarah gisteren heeft al haar vrienden een klein cadeautje gegeven. (*condition*: ungrammatical)
3. Geert zegt dit jaar nog niets te hebben gehoord van zijn vrienden uit Frankrijk. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* in correct position)
4. Saskia zegt afgelopen weekend weer veel in haar tuin hebben kunnen te werken. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* displaced)

5. Kasper hoopt vanaf deze maand vaker naar de volleybaltraining kunnen te komen. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* displaced)

Practice round II

1. Eva zegt dat haar vriend een nieuwe baan heeft gevonden. (*condition*: grammatical)

2. Sylvia morgen moet op tijd beginnen met werken. (*condition*: ungrammatical)

3. Kasper zegt nooit meer met zijn neef te willen voetballen. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* in correct position)

4. Thomas zegt vorige week weer lang op de trein hebben te moeten wachten. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* displaced)

5. Sarah zegt elke dag wel willen te gaan winkelen. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* displaced)

Practice round III

1. Peter vertelt dat hij graag op bezoek gaat bij zijn oma.(*condition*: grammatical)

2. Sarah volgend jaar wil graag een grote reis maken. (*condition*: ungrammatical)

3. Kasper zegt zo snel mogelijk klaar te willen zijn met opruimen. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* in correct position)

4. Geert zegt vorige week iedere dag langer hebben te moeten werken. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* displaced)

5. Thomas zegt de komende dagen zijn huis willen te schilderen. (*condition*: practice item with a verb selecting a *te*-infinitive and *te* displaced)

Fillers

1. Thomas maakt met plezier lunch voor zijn kinderen. (*condition*: grammatical)

2. Thomas belooft vanmiddag de tandarts te bellen. (*condition*: grammatical)

3. Thomas en Eva gaan vaak op zaterdag naar de markt. (*condition*: grammatical)

4. Eva zegt dat ze 's ochtends graag de krant leest. (*condition*: grammatical)

5. Eva heeft dit weekend met veel vriendinnen afgesproken. (*condition*: grammatical)

6. Sylvia bedenkt dat ze vanmiddag nog boodschappen moet doen. (*condition*: grammatical)

7. Sylvia belooft dat ze morgen de was zal doen. (*condition*: grammatical)

8. Sylvia hoopt dat ze volgende week minder kan werken. (*condition*: grammatical)

9. Robert praat niet graag over emotionele zaken. (*condition*: grammatical)

10. Robert zegt dat hij zaterdag gaat voetballen. (*condition*: grammatical)

11. Robert en zijn vrouw gaan in de zomer zijn oom in Frankrijk opzoeken. (*condition*: grammatical)

12. Jan hoopt dat zijn vriendin vanavond mee naar de bioscoop wil gaan. (*condition*: grammatical)

13. Jan zegt dat hij zelden bij zijn ouders langsgaat. (*condition*: grammatical)

14. Jan en Eva zijn erg gelukkig met hun nieuwe auto. (*condition*: grammatical)

15. Thomas kijkt uit naar zijn verjaardagsfeest. (*condition*: grammatical)

16. Thomas eet vaak meer dan goed voor hem is. (*condition*: grammatical)

17. Johan hoeft morgen waarschijnlijk niet te werken. (*condition*: control sentence with finite, third person singular *hoeft* 'need' in Verb Second position; *te* present)

18. Johan hoeft morgen waarschijnlijk niet werken. (*condition*: control sentence with finite, third person singular *hoeft* 'need' in Verb Second position; *te* absent)

19. Johan en Anne hoeven morgen waarschijnlijk niet te werken. (*condition*: control sentence with finite, third person plural *hoeven* 'need' in Verb Second position; *te* present)

20.

21. Johan en Anne hoeven morgen waarschijnlijk niet werken. (*condition*: control sentence with finite, third person plural *hoeven* 'need' in Verb Second position; *te* absent)

22. Johan zegt dat hij morgen waarschijnlijk niet hoeft te werken. (*condition*: control sentence with finite, third person singular *hoeft* 'need' in clause-final position; *te* present)

23. Johan zegt dat hij morgen waarschijnlijk niet hoeft werken. (*condition*: control sentence with finite, third person singular *hoeft* 'need' in clause-final position; *te* absent)

24. Johan en Anne zeggen dat ze morgen waarschijnlijk niet hoeven te werken. (*condition*: control sentence with finite, third person plural *hoeven* 'need' in clause-final position; *te* present)

25. Johan en Anne zeggen dat ze morgen waarschijnlijk niet hoeven werken. (*condition*: control sentence with finite, third person plural *hoeven* 'need' in clause-final position; *te* absent)

26. Eva zit weer de hele avond te praten. (*condition*: control sentence with finite, third person singular *zit* 'sit' in Verb Second position; *te* present)

27. Eva zit weer de hele avond praten. (*condition*: control sentence with finite, third person singular *zit* 'sit' in Verb Second position; *te* absent)

28. Johan en Eva zitten weer de hele avond te praten. (*condition*: control sentence with finite, third person plural *zitten* 'sit' in Verb Second position; *te* present)

29. Johan en Eva zitten weer de hele avond praten. (*condition*: control sentence with finite, third person plural *zitten* 'sit' in Verb Second position; *te* absent)

30. Anne vindt het vervelend dat Eva weer de hele avond zit te praten. (*condition*: control sentence with finite, third person singular *zit* 'sit' in clause-final position; *te* present)

31. Anne vindt het vervelend dat Eva weer de hele avond zit praten. (*condition*: control sentence with finite, third person singular *zit* 'sit' in clause-final position; *te* absent)

32. Anne vindt het vervelend dat Johan en Eva weer de hele avond zitten te praten. (*condition*: control sentence with finite, third person plural *zitten* 'sit' in clause-final position; *te* present)

33. Anne vindt het vervelend dat Johan en Eva weer de hele avond zitten praten. (*condition*: control sentence with finite, third person plural *zitten* 'sit' in clause-final position; *te* absent)

34. Johan heeft gisteren niet hoeven te werken. (*condition*: additional filler)

35. Johan heeft gisteren niet hoeven werken. (*condition*: additional filler)

36. Johan heeft gisteren niet hoeven te komen werken. (*condition*: additional filler)

37. Johan heeft gisteren niet hoeven komen werken. (*condition*: additional filler)

38. Voetballen heeft Koen gisteren niet willen. (*condition*: gramamtical)

39. Voetballen heeft Koen gisteren niet gewild. (*condition*: ungrammatical)

40. Uitslapen heb ik Anne gisteren niet laten. (*condition*: additional filler)

41. Uitslapen heb ik Anne gisteren niet gelaten. (*condition*: additional filler)

42. Praten zal Eva wel weer de hele avond hebben zitten. (*condition*: additional filler)

43. Praten zal Eva wel weer de hele avond te hebben zitten. (*condition*: additional filler)

44. Praten zal Eva wel weer de hele avond hebben te zitten. (*condition*: additional filler)

45. Praten zal Eva wel weer de hele avond hebben zitten te. (*condition*: additional filler)

46. Te praten zal Eva wel weer de hele avond hebben zitten. (*condition*: additional filler)

47. Jan hoopt volgend jaar met zijn nieuwe baan meer geld te verdienen. (*condition*: grammatical)

48. Jan hoopt volgend jaar met zijn nieuwe baan te verdienen meer geld. (*condition*: ungrammatical)

49. Sarah zegt dat haar zus dat boek heeft willen kopen. (*condition*: control sentence finite verb cluster in 123 order)

50. Sarah zegt dat haar zus dat boek willen heeft kopen. (*condition*: control sentence finite verb cluster in 213 order)

51. Thomas zegt dat zijn collega niet heeft willen luisteren naar zijn advies. (*condition*: grammatical)

52. Thomas zegt dat zijn collega niet heeft gewild luisteren naar zijn advies. (*condition*: ungrammatical)

53. Eva zegt dat haar vrienden het huis toch hebben kunnen gekocht. (*condition*: ungrammatical)

54. Eva zegt dat haar vrienden het huis toch hebben kunnen kopen. (*condition*: grammatical)

55. Eva zegt dat ze haar ouders gisteren helaas niet heeft kunnen gezien. (*condition*: ungrammatical)

56. Eva zegt dat ze haar ouders gisteren helaas niet heeft kunnen zien. (*condition*: grammatical)

57. Anne zegt dat ze haar oma gisteren lang is blijven helpen. (*condition*: control sentence finite verb cluster in 123 order)

58. Anne zegt dat ze haar oma gisteren lang is helpen blijven. (*condition*: control sentence finite verb cluster in 132 order)

59. Anne zegt dat ze haar oma gisteren lang blijven helpen is. (*condition*: control sentence finite verb cluster in 231 order)

60. Anne zegt dat ze haar oma gisteren lang blijven is helpen. (*condition*: control sentence finite verb cluster in 213 order)

61. Anne zegt dat ze haar oma gisteren lang helpen is blijven. (*condition*: control sentence finite verb cluster in 312 order)

62. Anne zegt dat ze haar oma gisteren lang helpen blijven is. (*condition*: control sentence finite verb cluster in 321 order)

63. Peter zal gisteren wel weer zijn bureau hebben zitten op te ruimen. (*condition*: additional filler)

64. Peter zal gisteren wel weer zijn bureau hebben zitten opruimen. (*condition*: additional filler)

65. Peter zal gisteren wel weer zijn bureau hebben zitten te opruimen. (*condition*: additional filler)

66. Johan wist niet wat te doen nadat zijn contract was afgelopen. (*condition*: additional filler)

67. Johan wist niet wat doen nadat zijn contract was afgelopen. (*condition*: additional filler)

68. Johan wist niet wat gedaan nadat zijn contract was afgelopen. (*condition*: additional filler)

69. Sarah zal vanavond toch langer te moeten blijven werken. (*condition*: control sentence verb cluster with unselected *te*)

70. Sarah zal vanavond toch langer moeten te blijven werken. (*condition*: control sentence verb cluster with unselected *te*)

71. Sarah zal vanavond toch langer moeten blijven te werken. (*condition*: control sentence verb cluster with unselected *te*)

Test items

1. Anne zegt haar oma gisteren lang te zijn blijven helpen. (*condition*: *te*-V1-V2-V3)

2. Anne zegt haar oma gisteren lang zijn te blijven helpen. (*condition*: V1-*te*-V2-V3)

3. Anne zegt haar oma gisteren lang zijn blijven te helpen. (*condition*: V1-V2-*te*-V3)

4. Anne zegt haar oma gisteren lang zijn blijven helpen. (*condition*: V1-V2-V3)

5. Anne zegt haar oma gisteren lang te zijn helpen blijven. (*condition*: *te*-V1-V3-V2)

6. Anne zegt haar oma gisteren lang zijn te helpen blijven. (*condition*: V1-*te*-V3-V2)

7. Anne zegt haar oma gisteren lang zijn helpen te blijven. (*condition*: V1-V3-*te*-V2)

8. Anne zegt haar oma gisteren lang zijn helpen blijven. (*condition*: V1-V3-V2)

9. Anne zegt haar oma gisteren lang te blijven zijn helpen. (*condition*: *te*-V2-V1-V3)

10. Anne zegt haar oma gisteren lang blijven te zijn helpen. (*condition*: V2-*te*-V1-V3)

11. Anne zegt haar oma gisteren lang blijven zijn te helpen. (*condition*: V2-V1-*te*-V3)

12. Anne zegt haar oma gisteren lang blijven zijn helpen. (*condition*: V2-V1-V3)

13. Anne zegt haar oma gisteren lang te blijven helpen zijn. (*condition*: *te*-V2-V3-V1)

14. Anne zegt haar oma gisteren lang blijven te helpen zijn. (*condition*: V2-*te*-V3-V1)

15. Anne zegt haar oma gisteren lang blijven helpen te zijn. (*condition*: V2-V3-*te*-V1)

16. Anne zegt haar oma gisteren lang blijven helpen zijn. (*condition*: V2-V3-V1)

17. Anne zegt haar oma gisteren lang te helpen zijn blijven. (*condition*: *te*-V3-V1-V2)

18. Anne zegt haar oma gisteren lang helpen te zijn blijven. (*condition*: V3-*te*-V1-V2)

19. Anne zegt haar oma gisteren lang helpen zijn te blijven. (*condition*: V3-V1-*te*-V2)

20. Anne zegt haar oma gisteren lang helpen zijn blijven. (*condition*: V3-V1-V2)

21. Anne zegt haar oma gisteren lang te helpen blijven zijn. (*condition*: *te*-V3-V2-V1)

22. Anne zegt haar oma gisteren lang helpen te blijven zijn. (*condition*: V3-*te*-V2-V1)

23. Anne zegt haar oma gisteren lang helpen blijven te zijn. (*condition*: V3-V2-*te*-V1)

24. Anne zegt haar oma gisteren lang helpen blijven zijn. (*condition*: V3-V2-V1)

25. Johan zal morgen waarschijnlijk niet te hoeven gaan werken. (*condition*: *te*-V1-V2-V3)

26. Johan zal morgen waarschijnlijk niet hoeven te gaan werken. (*condition*: V1-*te*-V2-V3)

27. Johan zal morgen waarschijnlijk niet hoeven gaan te werken. (*condition*: V1-V2-*te*-V3)

28. Johan zal morgen waarschijnlijk niet hoeven gaan werken. (*condition*: V1-V2-V3)

29. Johan zal morgen waarschijnlijk niet te hoeven werken gaan. (*condition*: *te*-V1-V3-V2)

30. Johan zal morgen waarschijnlijk niet hoeven te werken gaan. (*condition*: V1-*te*-V3-V2)

31. Johan zal morgen waarschijnlijk niet hoeven werken te gaan. (*condition*: V1-V3-*te*-V2)

32. Johan zal morgen waarschijnlijk niet hoeven werken gaan. (*condition*: V1-V3-V2)

33. Johan zal morgen waarschijnlijk niet te gaan hoeven werken. (*condition*: *te*-V2-V1-V3)

34. Johan zal morgen waarschijnlijk niet gaan te hoeven werken. (*condition*: V2-*te*-V1-V3)

35. Johan zal morgen waarschijnlijk niet gaan hoeven te werken. (*condition*: V2-V1-*te*-V3)

36. Johan zal morgen waarschijnlijk niet gaan hoeven werken. (*condition*: V2-V1-V3)

37. Johan zal morgen waarschijnlijk niet te gaan werken hoeven. (*condition*: *te*-V2-V3-V1)

38. Johan zal morgen waarschijnlijk niet gaan te werken hoeven. (*condition*: V2-*te*-V3-V1)

39. Johan zal morgen waarschijnlijk niet gaan werken te hoeven. (*condition*: V2-V3-*te*-V1)

40. Johan zal morgen waarschijnlijk niet gaan werken hoeven. (*condition*: V2-V3-V1)

41. Johan zal morgen waarschijnlijk niet te werken hoeven gaan. (*condition*: *te*-V3-V1-V2)

42. Johan zal morgen waarschijnlijk niet werken te hoeven gaan. (*condition*: V3-*te*-V1-V2)

43. Johan zal morgen waarschijnlijk niet werken hoeven te gaan. (*condition*: V3-V1-*te*-V2)

44. Johan zal morgen waarschijnlijk niet werken hoeven gaan. (*condition*: V3-V1-V2)

45. Johan zal morgen waarschijnlijk niet te werken gaan hoeven. (*condition*: *te*-V3-V2-V1)

46. Johan zal morgen waarschijnlijk niet werken te gaan hoeven. (*condition*: V3-*te*-V2-V1)

47. Johan zal morgen waarschijnlijk niet werken gaan te hoeven. (*condition*: V3-V2-*te*-V1)

48. Johan zal morgen waarschijnlijk niet werken gaan hoeven. (*condition*: V3-V2-V1)

Background questions (in Dutch)

Hartelijk dank voor het invullen van de enquête!

Tot slot willen we u nog vragen enkele persoonlijke gegevens in te vullen. Deze gegevens kunnen ons helpen bij het interpreteren van de ingevulde enquêtes. Wij willen nogmaals benadrukken dat deze gegevens anoniem zijn, en dat ze enkel binnen het kader van dit onderzoek gebruikt zullen worden.

1. Wat is/zijn uw moederta(a)l(en)?
2. Spreekt u een dialect? Zo ja welk:
3. Wat is uw geslacht? (mannelijk, vrouwelijk, anders)
4. Wat is uw leeftijd?
5. Wat is uw hoogst behaalde opleidingsniveau? (primair, secundair, MBO, hbo/hogeschool, universiteit)
6. In welk land bent u opgegroeid? (Nederland, België, anders)
7. In welke plaats bent u opgegroeid?
8. Tot welke leeftijd heeft u daar gewoond?
9. Heeft u ooit voor een tijd in het buitenland gewoond? Zo ja, hoelang: en waar:
10. Wat is uw huidige woonplaats? Sinds wanneer woont u daar?
11. Werkt u verder dan 10 kilometer van uw woonplaats af? Zo ja, hoeveel kilometer?

Appendix C

Instruction text for annotators (in English)

Dear annotator.

Thank you very much for participating in this study!

Could you please indicate for each sentence whether you would translate the verbal phrase as given in bold as ('... ' is the position of the main verb of the sentence):

1. 'went and ...', 'has/have gone and ...' or an equivalent construction with 'go and ...'
2. 'was ...-ing', 'has/have been ...-ing' or an equivalent progressive construction with 'be + -ing'
3. You are not sure

For example, the following sentence would be translated with a 'go and ...' construction, so option A would apply:

(1) 'Hulle het die boek **loop en verloor**!'

Translation by a native Afrikaans speaker:

(2) 'They **went and lost** the book!'

Whereas the following sentence would be translated with a progressive 'be ...-ing' construction, so option B would apply:

(3) 'Maar nee, nou **loop en beskinder** julle die arme man agter sy rug.'

Translation by a native Afrikaans speaker:

(4) ‘But no, now you **are gossiping** about the poor man behind his back.’

Note that these sentences come from a corpus with a variety of text genres and registers, so the sentences might not be as you would formulate them yourself, or might contain spelling mistakes/typos et cetera.

If you have any questions, please let me know.

Thanks again!

Appendix D

D1: Instruction text of the Afrikaans questionnaire study (in Afrikaans)

Introductie
Beste deelnemer,

Baie dankie vir u deelname aan hierdie studie, wat deel uitmaak van ‘n PhD-projek by die Katolieke Universiteit Leuven in België. U deelname geskied geheel en al anoniem, en die data wat in die konteks van die studie ingesamel word, sal slegs vir die doeleindes van die projek aangewend word. Voordat ons met die vraelys begin, wil ons u ‘n paar instruksies gee wat verband hou met hoe u die vraelys moet invul.

Instructie
Moedertaalsprekers van ‘n taal het dikwels ‘n baie sterk “gevoel” oor wat wel en nie tel as ‘n moontlike of aanvaarbare sin in hulle moedertaal. Dit is waar ten spyte van die feit dat hulle nooit spesifieke reëls geleer het om “reg te praat” nie. Wat moedertaalsprekers “maar net weet” sluit in hoe mens ‘n sin moet uitspreek, of die sin gepas is, gegewe die konteks, en ook of die struktuur van die sin korrek is.

Vir hierdie studie vra ons u om asseblief ‘n stel van 75 sinne te bekyk. Wat ons wil hê is dat u die sinne een vir een bekyk, en dat u vir elkeen aandui of die sin ‘n moontlike of onmoontlike sin is in die Afrikaans wat in u direkte omgewing gepraat word, d.w.s. klink die sin soos iets wat u vriende en familie en/of die mense in u tuisdorp of -stad sou gebruik? Probeer om op die sinstruktuur te fokus, en probeer om nie te veel te dink nie aan hoe waarskynlik dit sou wees dat u self die betrokke sin in ’n gesprek sou gebruik.

Vir van die sinne sal dit dalk vir u minder duidelik wees of u met ‘n moontlike of onmoontlike sin te doen het. U mag, byvoorbeeld, voel dat ‘n sin in beginsel moontlik is, maar dat dit nietemin vir u ietwat onnatuurlik of vreemd voorkom. In sulke gevalle sal u dit waarskynlik nuttig vind dat ons in hierdie studie 'n skaal gebruik wat van 1 tot 5 loop. Vir elke sin sal ons u vra:

Is hierdie sin 'n moontlike sin in Afrikaans soos dit in u direkte omgewing gepraat word?

As die antwoord op hierdie vraag 'beslis' is, ken u aan die sin 'n 5 toe. As die antwoord 'beslis nie' is, ken u 'n 1 toe. U kan ook van 4, 3 en 2 gebruik maak, met 4 vir sinne wat redelik goed, maar nie heeltemal perfek klink nie, en 2 vir dié wat redelik sleg, maar ook nie heeltemal onmoontlik klink; 3 kan u gebruik as die sin vir u middelmatig goed/sleg klink. As u regtig nie 'n gevoel oor 'n betrokke sin het nie, kan u op 'Ek weet nie.' kliek. Probeer egter om hierdie opsie so min soos moontlik te gebruik. As u wil uitbrei op 'n antwoord wat u gegee het, kan u dit doen deur u gedagtes aan te teken in die leë boks onderaan die betrokke sin.

LW: By sommige sinne sal daar 'n "ekstra" vraag oor die betekenis van die sin wees. Hierdie vraag moet u asseblief weer eens op grond van die 1-5 skaal beantwoord.

Lees elke sin hardop en kliek dan op u antwoord (1-5 of 'Ek weet nie.'). U eerste reaksie op elke sin is vir ons die belangrikste. Daar is geen regte of verkeerde antwoorde nie; wat saakmaak is u gevoel oor of 'n sin moontlik is of nie.

Die invul van die vraelys behoort ongeveer 20 minute van u tyd in beslag te neem. Soms sal die sinne wat u moet beoordeel baie eenders lyk. Ons wil u dus vra om asseblief nie te vergeet nie om elke sin hardop te sê voordat u die betrokke sin beoordeel. Die invul van hierdie vraelys is waarskynlik nie die opwindendste taak wat u nog onderneem het nie, maar u antwoorde is in die konteks van hierdie studie uiters belangrik. Ons sal dit dus hoog op prys stel as u die hele vraelys kan invul.

Dankie weer eens vir u deelname aan hierdie studie!

D2: List of practice items, fillers, test items and background questions (in Afrikaans)

Practice items

1. Elke Saterdag ontmoet die groepie vriende vir koffie langs die rivier. (*condition*: completely grammatical)
2. Gisteraand daardie studente het tot baie laat gedans. (*condition*: completely ungrammatical)

3. In die dorp gister was daar 'n geweldige protesoptog. (*condition*: medium grammatical)

4. Hulle het alles verloor in die oorstromings; hulle het nou letterlik niks oor. (*condition*: medium grammatical)

Fillers

1. Watter films stem hy in om te kyk en watter films weier hy om te?
2. Watter films stem hy in om te kyk en watter films weier hy om te doen?
3. Hoeveel woorde moet sy skryf en hoeveel (woorde) kan sy?
4. Hoeveel woorde moet sy skryf en hoeveel (woorde) kan sy doen?
5. Hy word gestraf en sy broer word ook.
6. Jan het 'n boek geskryf en Marie het ook.
7. Jan het 'n boek geskryf en Marie het ook gedoen.
8. Skryf doen hy die boek, maar hy wil dit nie publiseer nie.
9. Skryf kan ek daardie boek, maar ek kan dit nooit publiseer nie.
10. Die boek skryf sal ek waarskynlik, maar ek sal dit beslis nie publiseer nie.
11. Slaap sal sy waarskynlik.
12. Slaap sal sy waarskynlik slaap.

Test items

1. Eva sê dat Pieter die laaste tyd die hele middag loop en swem het. (*condition*: *loop*-PC with incompatible lexical verb, bare form of *loop*)

2. Eva sê dat Pieter die laaste tyd die hele middag geloop en swem het. (*condition*: *loop*-PC with incompatible lexical verb, *geloop*)

3. Sara sê dat Thomas die laaste tyd die hele middag sit en swem het. (*condition*: *sit*-PC with incompatible lexical verb, bare form of *sit*)

4. Sara sê dat Thomas die laaste tyd die hele middag gesit en swem het. (*condition*: *sit*-PC with incompatible lexical verb, *gesit*)

5. Laura sê dat Pieter die laaste tyd die hele middag staan en swem het. (*condition*: *staan*-PC with incompatible lexical verb, bare form of *staan*)

6. Laura sê dat Pieter die laaste tyd die hele middag gestaan en swem het. (*condition*: *staan*-PC with incompatible lexical verb, *gestaan*)

7. Ann sê dat Thomas die laaste tyd die hele middag lê en swem het. (*condition*: *lê*-PC with incompatible lexical verb, bare form of *lê*)

8. Ann sê dat Thomas die laaste tyd die hele middag gelê en swem het. (*condition*: *lê*-PC with incompatible lexical verb, *gelê*)

9. Steve sê dat Cornelia gisteraand baie loop en praat het. (*condition*: *loop*-PC progressive, *en* present, bare form of *loop*)

10. Steve sê dat Cornelia gisteraand baie geloop en praat het. (*condition*: *loop*-PC progressive, *en* present, *geloop*)

11. Steve sê dat Cornelia gisteraand baie loop praat het. (*condition*: *loop*-PC progressive, *en* absent, bare form of *loop*)

12. Steve sê dat Cornelia gisteraand baie geloop praat het. (*condition*: *loop*-PC progressive, *en* absent, *geloop*)

13. Paul sê dat Lisa verlede week 'n splinternuwe motor loop en koop het. (*condition*: *loop*-PC andative, *en* present, bare form of *loop*)

14. Paul sê dat Lisa verlede week 'n splinternuwe motor geloop en koop het. (*condition*: *loop*-PC andative, *en* present, *geloop*)

15. Paul sê dat lisa verlede week 'n splinternuwe motor loop en gekoop het. (*condition*: *loop*-PC andative, *en* present, bare form of *loop*, *ge-* on lexical verb)

16. Paul sê dat Lisa verlede week 'n splinternuwe motor loop koop het. (*condition*: *loop*-PC andative, *en* absent, bare form of *loop*)

17. Paul sê dat Lisa verlede week 'n splinternuwe motor geloop koop het. (*condition*: *loop*-PC andative, *en* absent, *geloop*)

18. Paul sê dat Lisa verlede week 'n splinternuwe motor loop gekoop het. (*condition*: *loop*-PC andative, *en* absent, bare form of *loop*, *ge-* on lexical verb)

19. Simon sê dat Thomas die hele middag sit en lees het. (*condition*: *sit*-PC, *en* present, bare form of *sit*)

20. Simon sê dat Thomas die hele middag gesit en lees het. (*condition*: *sit*-PC, *en* present, *gesit*)

21. Simon sê dat Thomas die hele middag sit en gelees het. (*condition*: *sit*-PC, *en* present, bare form of *sit*, *ge-* on lexical verb)

22. Simon sê dat Thomas die hele middag sit lees het. (*condition*: *sit*-PC, *en* absent, bare form of *sit*)

23. Simon sê dat Thomas die hele middag gesit lees het. (*condition*: *sit*-PC, *en* absent, *gesit*)

24. Simon sê dat Thomas die hele middag sit gelees het. (*condition*: *sit*-PC, *en* absent, bare form of *sit*, *ge*- on lexical verb)

25. Susan sê dat Elsa vir ure met haar ma op die telefoon staan en praat het. (*condition*: *staan*-PC, *en* present, bare form of *staan*)

26. Susan sê dat Elsa vir ure met haar ma op die telefoon gestaan en praat het. (*condition*: *staan*-PC, *en* present, *gestaan*)

27. Susan sê dat Elsa vir ure met haar ma op die telefoon staan praat het. (*condition*: *staan*-PC, *en* absent, bare form of *staan*)

28. Susan sê dat Elsa vir ure met haar ma op die telefoon gestaan praat het. (*condition*: *staan*-PC, *en* absent, *gestaan*)

29. Eric sê dat Michael die hele naweek lê en slaap het. (*condition*: *lê*-PC, *en* present, bare form of *lê*)

30. Eric sê dat Michael die hele naweek gelê en slaap het. (*condition*: *lê*-PC, *en* present, *gelê*)

31. Eric sê dat Michael die hele naweek lê slaap het. (*condition*: *lê*-PC, *en* absent, bare form of *lê*)

32. Eric sê dat Michael die hele naweek gelê slaap het. (*condition*: *lê*-PC, *en* absent, *gelê*)

33. Hoekom loop Jan heeldag piesangs en eet? (*condition*: normal V2 *loop*-PC, *en* present)

    - *If accepted*: Is hy besig om te loop? (Ja, nee, onduidlik)

34. Hoekom loop en eet Jan heeldag piesangs? (*condition*: quirky V2 *loop*-PC, *en* present)

    - *If accepted*: Is hy besig om te loop? (Ja, nee, onduidlik)

35. Hoekom loop Jan heeldag piesangs eet? (*condition*: normal V2 *loop*-PC, *en* absent)

    - *If accepted*: Is hy besig om te loop? (Ja, nee, onduidlik)

36. Hoekom loop eet Jan heeldag piesangs? (*condition*: quirky V2 *loop*-PC, *en* absent)

    - *If accepted*: Is hy besig om te loop? (Ja, nee, onduidlik)

37. Hoekom sit Lisa heeldag die koerant en lees?? (*condition*: normal V2 *sit*-PC, *en* present)
    - *If accepted*: Is hy besig om te sit? (Ja, nee, onduidlik)
38. Hoekom sit en lees Lisa heeldag die koerant? (*condition*: quirky V2 *sit*-PC, *en* present)
    - *If accepted*: Is hy besig om te sit? (Ja, nee, onduidlik)
39. Hoekom sit Lisa heeldag die koerant lees? (*condition*: normal V2 *sit*-PC, *en* absent)
    - *If accepted*: Is hy besig om te sit? (Ja, nee, onduidlik)
40. Hoekom sit lees Lisa heeldag die koerant? (*condition*: quirky V2 *sit*-PC, *en* absent)
    - *If accepted*: Is hy besig om te sit? (Ja, nee, onduidlik)
41. Hoekom staan Thomas heeldag sy ken en vryf? (*condition*: normal V2 *staan*-PC, *en* present)
    - *If accepted*: Is hy besig om te staan? (Ja, nee, onduidlik)
42. Hoekom staan en vryf Thomas heeldag sy ken? (*condition*: quirky V2 *staan*-PC, *en* present)
    - *If accepted*: Is hy besig om te staan? (Ja, nee, onduidlik)
43. Hoekom staan Thomas heeldag sy ken vryf? (*condition*: normal V2 *staan*-PC, *en* absent)
    - *If accepted*: Is hy besig om te staan? (Ja, nee, onduidlik)
44. Hoekom staan vryf Thomas heeldag sy ken? (*condition*: quirky V2 *staan*-PC, *en* absent)
    - *If accepted*: Is hy besig om te staan? (Ja, nee, onduidlik)
45. Hoekom lê Mark heeldag na die wolke en kyk? (*condition*: normal V2 *lê*-PC, *en* present)
    - *If accepted*: Is hy besig om te lê? (Ja, nee, onduidlik)
46. Hoekom lê en kyk Mark heeldag na die wolke? (*condition*: quirky V2 *lê*-PC, *en* present)
    - *If accepted*: Is hy besig om te lê? (Ja, nee, onduidlik)
47. Hoekom lê Mark heeldag na die wolke kyk? (*condition*: normal V2 *lê*-PC, *en* absent)

- *If accepted*: Is hy besig om te lê? (Ja, nee, onduidlik)

48. Hoekom lê kyk Mark heeldag na die wolke? (*condition*: quirky V2 *lê*-PC, *en* absent)

    - *If accepted*: Is hy besig om te lê? (Ja, nee, onduidlik)

49. Elsa het daaroor loop te dink. (*condition*: *loop* with *te*-complement)

50. Elsa het daaroor sit te dink. (*condition*: *sit* with *te*-complement)

51. Elsa het daar lê te rus. (*condition*: *lê* with *te*-complement)

52. Elsa het daar staan te wag. (*condition*: *staan* with *te*-complement)

53. Lisa sê dat Frank die CD gestap en koop het. (*condition*: non-PC motion verb *stap* in PC-construction, *gestap*, *en* present)

54. Lisa sê dat Frank die CD stap en koop het. (*condition*: non-PC motion verb *stap* in PC-construction, bare form of *stap*, *en* present)

55. Lisa sê dat Frank die CD stap koop het. (*condition*: non-PC motion verb *stap* in PC-construction, bare form of *stap*, *en* absent)

56. Frank stap die CD en koop. (*condition*: non-PC motion verb *stap* in PC-construction, normal V2)

57. Frank stap en koop die CD. (*condition*: non-PC motion verb *stap* in PC-construction, quirky V2)

58. Theresa sê dat Daniel sy hoed vergeet het in die kombuis. (*condition*: past participle with verbal prefix without *ge-* (*vergeet*), *het* present)

59. Theresa sê dat Daniel sy hoed gevergeet het in die kombuis. (*condition*: past participle with verbal prefix with *ge-* (*gevergeet*), *het* present)

60. Theresa sê dat Daniel sy hoed gevergeet in die kombuis. (*condition*: past participle with verbal prefix with *ge-* (*gevergeet*), *het* absent)

61. Paul sê dat hy die waarheid bewys het vir sy broer. (*condition*: past participle with verbal prefix without *ge-* (*bewys*), *het* present)

62. Paul sê dat hy die waarheid gebewys het vir sy broer. (*condition*: past participle with verbal prefix with *ge-* (*gebewys*), *het* present)

63. Paul sê dat hy die waarheid gebewys vir sy broer. (*condition*: past participle with verbal prefix with *ge-* (*gebewys*), *het* absent)

Background questions

Baie dankie vir die invul van die vraelys!

Ten laaste wil ons u graag ‘n paar vrae oor u agtergrond vra. Die antwoorde sal vir ons baie nuttig wees wanneer ons die uitslae van die vraelys interpreteer. Onthou dat al hierdie inligting anoniem is en dat dit uitsluitlik vir die interpretasie van die vraelysinligting gebruik sal word.

- Wat is u geslag (manlik, vroulik, ander)?
- Hoe oud is u?
- Wat is u hoogste akademiese kwalifikasie (hoërskool, voorgraads, nagraads)?
- Waar is u gebore? Gee asseblief u geboortedorp/stad.
- Hoe lank het u in die plek van u geboorte gebly?
- Waar woon u tans?
- Hoe lank woon u al daar?
- Wat is u moedertaal/tale (d.w.s. watter tale praat u sedert u geboorte of van toe u baie klein was [tot ongeveer peuter-ouderdom])?
- In watter taal voel u op u gemaklikste?
- Deur medium van watter taal/tale het u skoolonderrig plaasgevind?
- As u na skool gestudeer het, wat was die voertaal vir u nagraadse studies?
- Was Afrikaans een van u hoërskoolvakke?
- Was Afrikaans een van u voor/nagraadse studievakke?
- Hoe gebruik u Afrikaans? Met vriende, familie, by die werk, in daaglikse aktiwiteite soos inkopies doen, pendel, ens.
- Het u al ooit vir langer as ’n jaar oorsee gewoon? Indien wel, vir hoe lank?
- Gee asseblief aan: *eens*, *oneens* of *onseker*
  1. Die Afrikaans wat mens vandag hoor, is meestal deurspek van woorde wat van ander tale afkomstig is; bittermin mense praat ‘suiwer’ Afrikaans.
  2. Afrikaans is vandag ’n baie meer inklusiewe taal as wat hy lank in sy geskiedenis was.

3. Die meeste moderne Afrikaanssprekendes steur hulle nie meer aan taalreëls nie.
4. Dit is belangrik om te verseker dat Afrikaans hoorbaar en sienbaar is; dit moet nie net in 'n huistaal verander nie.

Bibliography

Abels, Klaus. 2016. The fundamental left-right asymmetry in the Germanic verb cluster. *The Journal of Comparative Germanic Linguistics* 19:179–220.

Acedo-Matellán, Víctor, and Cristina Real-Puigdollers. 2019. Roots into functional nodes: Exploring locality and semi-lexicality. *The Linguistic Review* 1–26.

Adger, David. 2006. Combinatorial variability. *Journal of Linguistics* 42:503–530.

Adger, David. 2014. Variability and Gramatical Architecture. In *Linguistic Variation in the Minimalist Framework*, ed. Carme M. Picallo, 179–196. New York: Oxford University Press.

Adger, David. 2016. Language Variability in Syntactic Theory. In *Rethinking Parameters*, ed. L. Eguren, O. Fernadez-Soriano, and A Mendikoetxea, 49–63. New York: Oxford University Press.

Adger, David, and Jennifer Smith. 2005. Variation and the Minimalist Program. In *Syntax and variation. Reconciling the biological and the social*, ed. Leonie Cornips and Karen P. Corrigan. John Benjamins Publishing Company.

Adger, David, and Jennifer Smith. 2010. Variation in agreement: a lexical feature-based approach. *Lingua* 120:1109–1134.

Aelbrecht, Lobke. 2009. You have the right to remain silent: The syntactic licensing of ellipsis. Doctoral Dissertation, Catholic University of Brussels.

Aelbrecht, Lobke. 2010. *The syntactic licensing of ellipsis*. John Benjamins Publishing Company.

Alexiadou, Artemis, Liliane Haegeman, and Melita Stavrou. 2008. *Noun Phrase in the generative perspective*. Berlin: Mouton de Gruyter.

Ambridge, B., J. Pine, and Lieven. E. 2013. Child language acquisition: Why Universal Grammar doesn't help. *Language* 90:53–90.

Augustinus, Liesbeth, and Peter Dirix. 2013. The IPP-effect in afrikaans: a corpus analysis. In *Proceedings of the 19th Nordic Conference of Computational Linguistics*, 213–225.

Augustinus, Liesbeth, and Peter Dirix. 2019. Gaan haal die sjampanje! een corpusanalyse van complex initials. Paper presented at the 6th Gentse Colloquium over het Afrikaans.

Bader, Thomas. 1995. Missing and misplaced z' in Bernese Swiss German. In *Topics in Swiss German syntax*, ed. Zvi Penner, 19–28. Bern: Peter Lang.

Barbiers, Sjef. 2014. Syntactic doubling and deletion as a source of variation. In *Linguistic Variation in the Minimalist Framework*, 197–223. New York: Oxford University Press.

Barbiers, Sjef, Johan van der Auwera, Hans Bennis, Eefje Boef, Gunther De Vogelaer, and Margreet van der Ham. 2008. *Syntactische atlas van de Nederlandse dialecten. Deel II*. Amsterdam: Amsterdam University Press.

Barbiers, Sjef, and Hans Bennis. 2010. De plaats van het werkwoord in zuid en noord. In *Voor Magda. Artikelen voor Magda Devos bij haar afscheid van de Universiteit Gent*, ed. Johan De Caluwe and Jacques Van Keymeulen, 25–42. Gent: Academia.

Barbiers, Sjef, Hans Bennis, and Lotte Dros-Hendriks. 2018. Merging verb cluster variation. *Linguistic Variation* 18:144–196.

Bennis, Hans. 2000. *Syntaxis van het Nederlands*. Amsterdam: Amsterdam University Press.

Bennis, Hans, and Teun Hoekstra. 1989. *Generatieve grammatica*. Dordrecht: Foris.

Beukema, F., and M. Den Dikken. 1989. The position of the infinitival marker in the germanic languages. In *Sentential complementation and the lexicon. studies in honour of wim de geest*, ed. D. Jaspers, W. Klooster, Y. Putseys, and P. Seuren, 75–75. Dordrecht: Foris.

Biberauer, Theresa. 2011. In defence of lexico-centric parametric variation: two 3rd factor-constrained case studies. Paper presented at the Workshop in Formal Grammar and Syntactic Variation: Rethinking Parameters.

Biberauer, Theresa. 2016. Three factors and modern generative diachronic syntax: Extensions and innovations ii: 'going beyond the input'. Handout as part of LOT summer school course. July 11-15 2016. Utrecht University, The Netherlands.

Biberauer, Theresa. 2017a. Factors 2 and 3: A principled approach. *Cambridge Occasional Papers in Linguistics* 10:38–65.

Biberauer, Theresa. 2017b. The Final-over-Final Condition and Particles. In *The Final-over-Final Condition: A Syntactic Universal*, ed. Michelle Sheehan, Theresa Biberauer, Ian Roberts, and Anders Holmberg, 187–297. MIT Press.

Biberauer, Theresa. 2017c. Pseudo-coordination in a hybrid system: (Parametric) insights from Afrikaans. Paper presented at the 1st Pseudo-Coordination and Multiple Agreement Constructions Meeting, May 2017.

Biberauer, Theresa. 2018. On the significance of the periphery: A 21st century generative perspective. Paper presented at WoSSP 15, June 2018.

Biberauer, Theresa. 2019a. Factors 2 and 3: towards a principled approach. *Catalan Journal of Linguistics special Issue* 1–44.

Biberauer, Theresa. 2019b. Grammaticalisation and pragmaticalisation have different entry points: evidence from pseudo-coordination. Paper presented at the 2nd Pseudo-Coordination and Multiple Agreement Constructions Meeting, March 2019.

Biberauer, Theresa, Cora Cavirani-Pots, Edoardo Cavirani, and Erin Pretorius. 2019. Auxiliary developments: A comparative perspective on Modern Afrikaans. Paper presented at the 34th Comparative Germanic Syntax Workshop, June 2019.

Biberauer, Theresa, and Marc Richards. 2006. When the grammar doesn't mind. In *Minimalist Essays*, ed. Cedric Boeckx, number 91 in Linguistics Today, 35–67. Amsterdam: John Benjamins.

Biberauer, Theresa, and Sten Vikner. 2017. Having the edge: A new perspective on pseudo-coordination in danish and afrikaans. In *A schrift to fest kyle johnson*, ed. Nicholas LaCara, Keir Moulton, and Anne-Michelle Tessier, 77–90. Benjamins.

Bjorkman, Bronwyn M. 2011. BE-ing Default: The Morphosyntax of Auxiliaries. Doctoral Dissertation, MIT.

Bjorkman, Bronwyn M., and Hedde Zeijlstra. 2014. Upward Agree is superior. Unpublished manuscript.

Bjorkman, Bronwyn M., and Hedde Zeijlstra. 2019. Checking up on (Phi-)agree. *Linguistic Inquiry* 50:527–569.

Blümel, Rudolf. 1914. *Einführing in die Syntax*. Heidelberg: C Winter.

Borer, Hagit. 2005a. *In Name Only: Structuring Sense*, volume I. Oxford: Oxford University Press.

Borer, Hagit. 2005b. *The Normal Course of Events: Structuring Sense*, volume II. Oxford: Oxford University Press.

Bošković, Željko. 2009. On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic Inquiry* 38:589–644.

Bosman, D. B. 1923. *Oor die ontstaan van Afrikaans*. Amsterdam: Swets & Zeitlinger.

Bouman, A. C. 1926. *Onderzoekingen over Afrikaanse sintaxis*. Cape Town: Nasionale Pers.

Breed, Adri. 2012. Die grammatikalisering van aspek in Afrikaans: 'n semantiese studie van perifrastiese progressiewe konstruktties. Doctoral Dissertation, North West University, Potchefstroom.

Breed, Adri. 2017. The subjective use of postural verb in afrikaans (i): Evolution from progressive to modal. *Stellenbosch Papers in Linguistics Plus* 1:1–21.

Breed, Adri, Frank Brisard, and Ben Verhoeven. 2017. Periphrastic progressive constructions in dutch and afrikaans: A contrastive analysis. *Journal of Germanic Linguistics* 4:305–379.

Cardinaletti, Anna, and Giuliana Giusti. 2001. "semi-lexical" motion verbs in Romance and Germanic. In *Semi-lexical categories: the function of content words and the content of function words*, ed. Norbert Corver and Henk van Riemsdijk, volume 59 of *Studies in Generative Grammar*, 371–414. Berlin: Mouton de Gruyter.

Cardinaletti, Anna, and Ur Shlonsky. 2004. Clitic positions and restructuring in Italian. *Linguistic Inquiry* 35:519–557.

Carstens, Vicki. 2015. Delayed valuation: A reanalysis of goal features, 'upward' complementizer agreement, and the mechanics of case. *Syntax* 19:1–42.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–156. MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in linguistics*, ed. Michael Kenstowicz, 1–52. Cambridge, Massachusetts: MIT Press.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: a cross-linguistic perspective*. New York: Oxford University Press.

Cinque, Guglielmo. 2001. "Restructuring" and the order of aspectual and root modal heads. In *Current studies in Italian syntax: Essays offered to Lorenzo Renzi*, ed. G. Cinque and G. Salvi, 137–155. Amsterdam: Elsevier.

Cinque, Guglielmo. 2003. The interaction of passive, causative and 'restructuring' in Romance. In *The syntax of Italian dialects*, ed. Christina Tortora, 50–66. Oxford University Press.

Cinque, Guglielmo. 2006. *Restructuring and functional heads*, volume 4 of *The Cartography of Syntactic Structures*. New York, New York: Oxford University Press.

Cinque, Guglielmo, and Luigi Rizzi. 2009. The cartography of syntactic structures. In *The Oxford handbook of syntactic structures*, ed. Bernd Heine and Heiko Narrog. Oxford: Oxford University Press.

Conradie, C. J. 1979. Die diachronie van die afrikaanse voltooide deelwoord. Doctoral Dissertation, University of the Witwatersrand.

Conradie, C. J. 1999. Preterite loss in early afrikaans. *Folia Linguistica* Special issue: Characteristic proprties of spoken vernaculars:19–38.

Conradie, C. J. 2007. The final stages of deflection: the case of afrikaans het 'have'. In *Historical Linguistics 2005. Selected papers from the 17th International Conference on Historical Linguistics*, ed. J. Salmons and S. Dubenion-Smith, 207–221. Amsterdam: Benjamins.

Conradie, C. J. 2012. The Dutch-Afrikaans participial prefix ge-: a case of degrammaticalisation? In *Historical linguistics 2009. selected papers from the 19th International Conference on Historical Linguistics*, ed. A. Van Kemenade and Nynke de Haas, 131–153. Philadelphia: John Benjamins.

Conradie, C. J. 2018. From auxiliary to inflectional ending? The case of Afrikaans het 'have'. Paper presented at the 51st SLE conference, August 2018.

van Craenenbroeck, Jeroen. 2010. *The syntax of ellipsis. Evidence from Dutch dialects*. New York: OUP.

Cremers, Crit. 1983. On two types of infinitival complementation. In *Linguistics categories: Auxiliaries and related puzzles*, ed. Frank Heny and Barry Richards, 169–221. Dordrecht: Reidel.

De Belder, Marijke. 2011. Roots and affixes. Eliminating lexical categories from syntax. Doctoral Dissertation, UiL-OTS/Utrecht University/HUBrussel.

De Belder, Marijke, and Jeroen Van Craenenbroeck. 2014. On Vocaburay Insertion. *Linguistic Analysis* 39:173–210.

De Belder, Marijke, and Jeroen Van Craenenbroeck. 2015. How to merge a root. *Linguistic Inquiry* 46:625–655.

De Klerk, W. J. 1968. Dialektiese verskeidenheid in Afrikaans. In *Afrikaans: sy aard en ontwikkeling*, ed. H. J. J. M. Van der Merwe, 219–239. Pretoria: Van Schaik.

De Rooij, J. J. 1981. 'het is gisteren beginnen te vriezen.' een syntactisch drieluikje voor vanacker. *Taal en Tongval* 33:23–30.

De Schutter, G. 2001. Een paar beschouwingen over het IPP-effect in het Afrikaans. *Taal & Tongval* 47:31–39.

De Vos, Mark Andrew. 2001. Afrikaans verb clusters: A functional-head analysis. Master's thesis, University of Tromsø, Norway.

De Vos, Mark Andrew. 2003. Past participles in Afrikaans dialects and Dutch. In *WCCFL 22 Proceedings*, ed. G Garding and M Tsujimura, 519–532. Somerville, Massachusetts: Cascadilla Press.

De Vos, Mark Andrew. 2005. The syntax of verbal pseudo-coordination in English and Afrikaans. Doctoral Dissertation, Utrecht, The Netherlands.

Den Besten, H., and H. Broekhuis. 1989. Woordvolgorde in de werkwoordelijke eindreeks. *GLOT* 12:79–137.

Den Besten, Hans. 1988. Universal-Grammatik und/oder Zweitsprachenerwerb: Der Fall Afrikaans. In *Beiträge zum 4. Essener Kolloquium über 'Sprachkontakt, Sprachwandel, Sprachwechsel, Sprachtod'*, ed. N. Boretzky, W. Enninger, and T. Stolz. Bochum: Studienverlag Dr. N. Brockmeyer.

Den Dikken, Marcel. 2002. Direct and parasitic polarity item licensing. *Journal of Comparative Germanic Linguistics* 5:35–66.

Den Dikken, Marcel, and C. Jan-Wouter Zwart. 1996. *Studies in comparative Germanic syntax ii*, chapter Very exceptional case marking, 85–108. Kluwer Academic Publishers.

Dik, Simon C. 1983. *Advances in Functional Grammar*, chapter Two Constraints on Relators and what they can dor for us, 267–298. Foris.

Dirix, Peter, Liesbeth Augustinus, and Frank Van Eynde. 2020. IPP in afrikaans: a corpus-based investigation and a comparison with Dutch and German. In *Konvergenz und Divergenz – sprachvergleichende Studien zum Deutschen*, ed. G. De Vogelaer, D. Koster, and T. Leuschner. Berlin/New York: De Gruyter.

Donaldson, B. 1993. *A grammar of afrikaans*. Berlin: Mouton de Gruyter.

Dreumel, Simon, and Peter-Arno Coppen. 2003. Surface analysis of the verbal cluster in dutch. *Linguistics* 41:51–81.

Dros-Hendriks, Lotte. 2018. Not another book on Verb Raising. Doctoral Dissertation, LOT Dissertation series.

Du Plessis Scholtz, J. 1963. Die volgorde van verbonde verbale vorme. In *Taalhistoriese Opstelle: Voorstudies tot 'n Geskiedenis van Afrikaans*, 162–168. J.L. van Schaik.

Du Toit, P.J. 1905. Afrikaansche studies. Doctoral Dissertation, Ghent University.

Gelderen, Elly van. 2011. *The linguistics cycle: Language change and the Language Faculty*. Oxford: Oxford University Press.

Gervain, J., and J. Mehler. 2010. Speech perception and language acquisition in the first year of life. *Annual Review of Psychology* .

Greenacre, Michael. 2007. *Correspondence analysis in practice*. Chapman & Hall, 2nd edition.

Grønn, Atle, and Arnim von Stechow. 2011. The temporal organisation of indicative conditionals. Unpublished manuscript.

Hachem, Mirjam. 2015. Multifunctionality: The internal and external syntax of D- and W-items in German and Dutch. Doctoral Dissertation, Utrecht Univeristy.

Haegeman, Liliane. 1995. Root infinitives, tense, and truncated structures in Dutch. *Language Acquisition* 4:205–255.

Haegeman, Liliane, and Terje Lohndal. 2010. Negative concord and (Multiple) Agree: a case study of West Flemish. *Linguistic Inquiry* 41:181–211.

Haeseryn, W., K. Romijn, G Geerts, J. de Rooij, and M.C. van den Toorn. 1997. *Algemene nederlandse spraakkunst. second edition*. Groningen and Deurne: Martinus Nijhoffuitgevers/Wolters Plantyn.

Hagemeijer, Tjerk. 2001. Underspecification and serial verb constructions. In *Semi-lexical categories: the function of content words and the content of function words*, ed. Norbert Corver and Henk van Riemsdijk, 415–454. Berlin/New York: Mouton de Gruyter.

Haider, Hubert. 2001. Heads and selection. In *Semi-lexical categories: the function of content words and the content of function words*, ed. Norbert Corver and Henk van Riemsdijk, 67–96. Berlin/New York: Mouton de Gruyter.

Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In *The view from building 20*, ed. Kenneth Hale and Samuel Jay Keyser, 111–176. Cambridge, Massachusetts: MIT Press.

Harley, Heidi, and Rolf Noyer. 1999. State-of-the-Article: Disttributed Morphology. *GLOT International* 4:3–9.

Haslinger, Irene, and Marjo van Koppen. 2002-2003. De verbale hendiadys als pseudocoördinatie. *Taal & Tongval* 15-16:102–122.

Heine, B., U. Claudi, and F. Huennemeyer. 1991. *Grammaticalization: A Conceptual Framework*. Chicago: Cicago University Press.

Henry, Alison. 2002. Variation and Syntactic Theory. In *The handbook of language variation and change*, ed. J. K. Chambers, P. Trudgill, and N. Schilling-Estes, 267–282. Blackwell Publishing.

Hinterhölzl, Roland. 2009. The IPP-effect, phrasal affixes and repair strategies in the syntax-morphology interface. *Linguistische Berichte* 218.

Hockett, Charles. 1958. A course in Modern Linguistics. Course taught in New York.

Holm, Hans Heinrich. 1988. *Pidgins and creoles. vol. 1: Theory and structure*. Cambridge: Cambridge University Press.

Holmberg, Anders. 2015. Verb Second. In *Handbook of Linguistics and Communication Science*, ed. T. Kiss and A. Alexiadou, volume 1, 342–383. Berlin/Boston: Mouton de Gruyer.

Hopper, P., and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Husson, Francois, Julie Josse, Sebastien Le, and Jeremy Mazet. 2014. *Factominer: Multivariate exploratory data analysis and data mining with r*. URL `http://CRAN.R-project.org/package=FactoMineR`, r package version 1.26.

Husson, François, Sébastien Lê, and Jérôme Pagès. 2011. *Exploratory multivariate analysis by example using R*. Boca Raton/London/New York: CRC Press.

IJbema, Aniek. 2001. Grammaticalization and infinitival complements in Dutch. Doctoral Dissertation, Leiden University.

Kempen, W. 1965. Die verbale hendiadis in Afrikaans. In *Dietse studies: Bundel aangebied aan prof. dr. j. du Plessis Scholtz*, ed. E. Lindenberg, E. H. Raidt, and J. A. Verhage, 77–97. Cape Town/Pretoria: Academia.

Klockmann, Heidi. 2017. The Design of Semi-lexicality. evidence from Case and Agreement in the Nominal Domain. Doctoral Dissertation, LOT Dissertation series.

Kocks, Denise. 1951. Die verbale hendiadis in Afrikaans. MA thesis, University of Stellenbosch.

Kroch, Anthony S. 2001. Syntactic Change. In *Handbook of Contemporary Syntactic Theory*, ed. Mark Baltin and Chris Collins, 699–729. Blackwell Publishing.

Kroes, H. 1982. Die frekwensiebepaling van die kernwoordeskat en sekere strukture van die afrikaanse spreektaal. Unpublished research report [Three corpora: Main, SABC and Spontaneous].

Kuteva, T. 1999. On 'sit'/'stand'/'lie' auxiliation. *Linguistics* 37:191–213.

Labov, W. 1966. *The Social Stratifiation of English in New York City*. Washington D.C.: Center for Applied Linguistics.

Labov, W. 1994. *Principles of Linguistic Change*, volume I: Internal Factors. Oxford: Blackwell Publishing.

Lehmann, C. 1985. Grammaticalisation, synchronic variation and diachronic change. *Lingua e stile* 20:303–318.

Lemmens, Maarten. 2005. Aspectual posture verb constructions in Dutch. *Journal of German Linguistics* 17:183–217.

Levshina, Natalia. 2015. *How to do linguistics with r. data exploration and statistical analysis.*. John Benjamins.

Leys, O. 1985. De konstruktie staat te+ infi nitief en verwante konstrukties. *Verslagen en Mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde* 13:265–277.

Lødrup, H. 2002. The syntactic structures of norwegian pseudocoordinations. *Studia Linguistica* 56:121–143.

Loubser, J.E. 1960. *Die saamgestelde verbale vorm van Nederlands na Afrikaans*. Groningen: Wolters.

Lowenstamm, Jean. 2014. Derivational affixes as roots: Phasal spell-out meet English stress shift. In *The Syntax of Roots and the Roots of Syntax*, ed. Artemis Alexiadou, Hagit Borer, and Florian Schäfer, 230–259. Oxford: Oxford University Press.

Makhudu, D. P. 1984. Is afrikaans a creole language? Master's thesis, Southern Illinois University, Ann Arbor: University Mcirofilms International.

Meurers, Walt Detmar. 2000. Lexical generalisations in the syntax of German non-finite constructions. Doctoral Dissertation, Universität Tübingen.

Meussen, A. E. 1943. Vier isotagmen. *Leuvense Bijdragen* XXXV:47–55.

Miller, Philip. 1991. Clitics and constituents in phrase structure grammar. Doctoral Dissertation, Utrecht University.

Müller, Gereon. 2009. Ergativity, accusativity, and the order of Merge and Agree. In *Explorations of Phase Theory: Features and Arguments*, 269–308. Grohmann, Kleanthes K.

Newman, John. 2002. A cross-linguistic overview of the posture verbs 'sit', 'stand' and 'lie'. In *The Linguistics of Sitting, Standing, and Lying*, ed. John Newman, 1–24. John Benjamins Publishing Company.

Nowenstein, I. 2014. Intra-speaker variation in subject Case: Icelandic. *University of Pennsylvania Working Papers in Linguistics* 20:Article 28.

Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste, and Ineke Schuurman. 2013. The construction of a 500-million-word reference corpus of contemporary written dutch. In *Essential speech and language technology for dutch: Results by the stevin-programme*, ed. Peter Spyns and Jan Odijk. Springer.

Pardoen, J. 1986. Werkwoordclustering in de voltooide tijd. In *Jaarboek voor de Neerlandistiek vii*, 49–76.

Pesetsky, David, and Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and clausal architecture*, ed. Simin Karimi, Vida Samiian, and Wendy K. Wilkins, 262–294. Amsterdam: John Benjamins.

Philip, Joy Naomi. 2012. Subordinating and Coordinating Linkers. Doctoral Dissertation, University College London, London.

Pinker, Steven. 1984. *Language Learnability and Language Development*. Cambridge, Massachusetts: Harvard University Press.

Poletto, Cecilia. 2000. *The higher functional field: Evidence from Northern Italian dialects*. Oxford University Press.

Ponelis, F. 1968. *Grondtrekke van die Afrikaanse Sintaksis*. Pretoria: J.L. van Schaik.

Ponelis, F. 1979. *Afrikaanse Sintaksis*. Pretoria: J.L. van Schaik, first edition edition.

Ponelis, F. 1993. *The Development of Afrikaans*. Frankfurt am Main: Die Deutsche Bibliothek.

Pots, Cora. 2017. Te-raising is clitic climbing. *Linguistics in the Netherlands* 34:127–141.

Preminger, Omer. 2015. Upwards and onwards: On the direction of valuation in phi-feature agreement. NYU Syntax Brown-Bag Talk handout.

R Core Team. 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

Rademeyer, J. H. 1938. *Kleurling-Afrikaans: Die taal van die Griekwas en Rehoboth-Basters*. Swets & Zeitlinger.

Raidt, E. H. 1983. *Einführung in Geschicthe und Structur des Afrikaans*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Ramchand, Gillian C. 2018. *Situations and Syntactic Structures*. MIT Press Monographs.

Richards, Marc. 2008. Tho kinds of variation in a Minimalist system. In *Varieties of Competition*, ed. Fabian Heck, Gereon Müller, and Jochen Trommer, volume 87, 133–162. Universität Leipzig.

Riemsdijk, Henk van. 1998a. Categorial feature magnetism: The endocentricity and distribution of projections. *The Journal of Comparative Germanic Linguistics* 2:1–48.

Riemsdijk, Henk van. 1998b. Head movement and adjacency. *Natural Language and Linguistic Theory* 16:633–678.

Rizzi, Luigi. 1978. A restructuring rule in italian syntax. In *Recent transformational studies in European languages*, ed. Samuel Jay Keyser, 113–158. Cambridge, Mass.: MIT Press.

Rizzi, Luigi. 2016. Labeling, maximality and the head – phrase distinction. *The Linguistic Review* 33:103–127.

Robbers, Karin. 1997. Non-finite verbal complements in Afrikaans. Doctoral Dissertation, University of Amsterdam.

Roberge, Paul. 1994. On the origins of the Afrikaans verbal hendiadys. *Stellenbosch Papers in Linguistics* 28:45–81.

Roberts, Ian, and Anna Roussou. 2003. *Syntactic change: A Minimalist approach to grammaticalisation*. Cambridge: Cambridge University Press.

Ross, Daniel. 2016. Going to surprise: the grammaticalization of itive as mirative. Online proceedings of Cognitive Linguistics in Wrocław Web Conference 2016. https://sites.google.com/site/coglingwroc2/.

Ross, John. 1967. Constraints on variables in syntax. Doctoral Dissertation, Massachusetts Institute of Technology.

Rutten, J. 1991. Infinitival complements and auxiliaries. Doctoral Dissertation, University of Amsterdam.

Salzmann, Martin. 2013. Rule ordering in verb cluster formation: On the extraposition paradox and the placement of the infinitival particle te/zu. In *Rule interaction in grammar*, ed. F. Heck and A. Assmann, 65–121. Leipzig: Linguistische Arbeitsberichte, University of Leipzig.

Salzmann, Martin. 2016. Displaced morphology in german: Evidence for post-syntactic morphology. In *Replicative Processes in Grammar*, ed. K Barnickel, M. Guzmán, J. Hein, S. Korsah, A. Murphy, L. Paschen, Z. Puskar, and J. Zaleska, volume 93, 401–446. Linguistische Arbeitsberichte, University of Leipzig.

Salzmann, Martin. 2019a. Displaced morphology in german verb clustes: an argument for post-syntactic morphology. *Journal of Comparative Germanic Linguistics* 22:1–53.

Salzmann, Martin. 2019b. On the limits of variation in Continental West-Germanic verb clusters: Evidence from VP-stranding, extraposition and displaced morphology for the existence of clusters with 213 order. *The Journal of Comparative Germanic Linguistics* 22:1–53.

Schallert, Oliver. 2019. A note on misplaced or wrongly attached zu 'to' in German. *Journal of Germanic Linguistics* .

Schmid, Tanja. 2005. *Infinitival syntax: Infinitivus pro participio as a repair strategy*. Amsterdam: John Benjamins.

Schütze, Carson T. 2016. *The emprical base of linguistics: Grammaticality judgments and linguistic methodology*. Berlin: Language Science Press.

Schuurman, Ineke., and Annet Wierenga. 1986. Syntactische nomen-incorportatie bij infinitieven en deelwoorden. In *Proeven van taalwetenschap. ter gelegenheid van het emeritaat van albert sassen*, ed. C. Hoppenbrouwers, J. Houtman, I. Schuurman, and F. Zwarts, 339–350. Groningen: Centrale Repro RUG.

Sharvit, Yael. 2003. Trying to be progressive: the extensionality of try. *Journal of Semantics* 20:403–445.

Smith, P. 2015. Feature Mismatches: Consequences for Syntax, Morphology, and Semantics. Doctoral Dissertation, University of Connecticut.

Song, C. 2019. On the Formal Flexibility of Syntactic Categories. Doctoral Dissertation, University of Cambridge.

South, Andy, Joe Scutt-Phillips, Barry Rowlingson, Roger Bivand, and Pru Foster. 2016. *Rworldmap: Mapping global data*, r pacakge version 1.3-4 edition.

Stavrou, Melita. 2003. Semi-lexical nouns, classifiers, and the interpretation(s) of the pseudopartitive construction. In *From NP to DP: The syntax and semantics of noun phrases*, ed. Y. D'hulst and M. Coene, volume I, 329–354. John Benjamins.

Stechow, Arnim von. 1990. Status government and coherence in German. In *Scrambling and barriers*, ed. Günther Grewendorf and Wolfgang Sternefeld, 143–198. Amsterdam: John Benjamins Publishing Company.

Stechow, Arnim von. 2003. Feature deletion under semantic binding: Tense, person, and mood under verbal quantifiers. In *Proceedings of the North East Linguistic Society 33*, ed. Makoto Kadowaki and Shigeto Kawahara, 379–403.

Sternefeld, Wolfgang. 1990. Scrambling and minimality. In *Scrambling and barriers*, ed. Günther Grewendorf and Wolfgang Sternefeld, 239–260. Amsterdam: John Benjamins Publishing Company.

Stoett, F. A. 1923. *Middelnederlandsche spraakkunst*. The Hague: Martinus Nijhoff, 3rd edition edition.

Stolz, T. 1986. *Gibt es das kreolische Srachwandelmodell? Vergleichende Grammatik des Negerholländischen*. Frankfurt am Main/Bern/New York: Peter Lang.

Tamminga, M., L. MacKenzie, and D. Embick. 2016. The dynamics of variation in individuals. *Linguistic Variation* 2:300–336.

Ter Beek, Janneke. 2008. Restructuring and infinitival complements in dutch. Doctoral Dissertation, LOT Dissertation series.

Thoms, G., D. Adger, C. Heycock, and J. Smith. 2019. Syntactic variation and auxiliary contraction: the surprising case of Scots. *Language* 95:421–455.

Tortora, Christina M. 2014. Addressing the problem of intra-speaker variation for parametric theory. In *Micro-syntactic variation in North-American English*, ed. Rafaella Zanuttini and Laurence R. Horn, 294–323. Oxford: Oxford University Press.

Traugott, E., and B Heine. 1991. *Approaches to Grammaticalization*. Amsterdam: John Benjamins.

Valkhoff, M. F. 1972. *New light on Afrikaans and 'Malayo-portuguese'*. Leuven: Éditions Peeters Imprimerie Orientaliste.

Van Craenenbroeck, Jeroen, and Marjo Van Koppen. 2017. Look! kaatje is still not heard sing a song (even without government). In *GLOW workshop What happened to government?*.

Van Craenenbroeck, Koppen Marjo van Bosch Antal van den, Jeroen. 2019. A quantitative-theoretical analysis of syntactic microvariation: Word order in Dutch verb clusters. *Language* 95:333–370.

Van Eynde, Frank, Liesbeth Augustinus, Ineke Schuurman, and Vincent Vandeghinste. 2016. Hebben of zijn bij IPP's. *Leuvense Bijdragen* 99-100:11–28.

Van Pottelberge, Jeroen. 2002. Nederlandse progressiefconstructies met werkwoorden van lichaamshouding. specificiteit en geschiedenis. *Nederlandse taalkunde* 7:142–174.

Van Rensburg, M.C.J. 1987. *Finale verslag van die groter navorskingsprojek: Gesproke Afrikaans: Die Afrikaans van die Richtersveld*. Bloemfontein: University of the Orange Free State.

Van Rensburg, M.C.J. 1989. Orange River Afrikaans: A stage in the pidgin/creole cycle. In *Wheels within wheels: Papers of the Duisburg Symposium on Pidgin and Creole Languages*, ed. M. Pütz and R. Dirven, 135–151. Frankfurt am Main/Bern/New York: Peter Lang.

Vanacker, Valeer Frits. 1969. Zuidnederlandse dialektmateriaal op de band en enkele dubbele werkwoordgroepen in te-positie. *Taal en Tongval* 21:239–244.

Van de Velde, Freek. 2017. Understanding grammar at the community level required a diachronic perspective. evidence from four case studies. *Nederlandse taalkunde* 22:47–74.

VivA. 2016. Korpusportaal. Available at http://viva-afrikaans.org. URL `http://viva-afrikaans.org`.

Vogel, Ralf. 2009. Skandal im Verbkomplex. betrachtungen zur scheinbar inkorrekten Morphologie in infiniten Verbkomplexen des Deutschen. *Zeitschrift für Sprachwissenschaft* 28:307–346.

Weinreich, U., W. Labov, and M. I. Herzog. 1968. *Empirical Foundations for a Theory of Language Change*. Austin: University of Texas Press.

Weisser, Philipp. 2015. *Derived Coordination. A Minimalist Perspective on Clause Chains, Converbs and Asymmetric Coordination*, volume 561 of *Linguistische Arbeiten*. De Gruyter.

Wiklund, Anna-Lena. 2007. *The Syntax of Tenselessness: Tense/Mood/Aspect-Agreeing Infinitivals*, volume 92 of *Studies in Generative Grammar*. Mouton de Gruyter.

Wiltschko, Martina. 2014. *The Universal Structure of Categories. towards a Formal Typology*. Cambridge: Cambridge University Press.

Wurmbrand, Susi. 2001. *Infinitives: Restructuring and clause structure*. Mouton de Gruyter.

Wurmbrand, Susi. 2004. West Germanic verb clusters: the empirical domain. In *Verb clusters. A study of Hungarian, German and Dutch*, ed. Katalin É. Kiss and Henk van Riemsdijk, 43–86. Amsterdam: John Benjamins.

Wurmbrand, Susi. 2012a. Parasitic participles: Evidence for the theory of verb clusters. *Taal en Tongval* 64:129–156.

Wurmbrand, Susi. 2012b. The syntax of valuation in auxiliary-participle constructions. In *Coyote Working Papers: Proceedings of the 20th West Coast Conference on Formal Linguistics*, ed. Jaehoon et al. Choi, 154–162. Tucson: University of Arizona.

Wurmbrand, Susi. 2015. Restructuring cross-linguistically. In *Proceedings of the North East Linguistic Society 45*, ed. T. Bui and D. Özyildiz, 227–240. Amherst, MA: GLSA.

Wurmbrand, Susi. 2017. Verb clusters, verb raising, and restructuring. In *The Blackwell Companion to Syntax. 2nd edition.*, ed. Martin Everaert and Henk van Riemsdijk. Hoboken: Wiley-Blackwell.

Wurmbrand, Susi, and Magdalena Lohninger. 2019. An implicational universal in complementation: Theoretical insights and empirical progress. In *Propositonal Arguments in Cross-Linguistic Research: Theoretical and Empirical Issues*, ed. Jutta M. Hartmann and Angelika Wöllstein. Berlin: Mouton de Gruyer.

Zeijlstra, Hedde. 2008. On the syntactic flexibility of formal features. In *The limits of syntactic variation*, ed. Theresa Biberauer, 143–173. Amsterdam: John Benjamins Publishing Company.

Zeijlstra, Hedde. 2012. There is only one way to agree. *The Linguistic Review* 29:441–483.

Zhang, N. 2007. Root merger in chinese compounds. *Studia Linguistica* 61:170–184.

Zoerner, Ed. 1999. One Coordinator for All. *Linguistic Analysis* 29:322–341.

Zwart, C. Jan-Wouter. 1993. Dutch syntax: A minimalist approach. Doctoral Dissertation, University of Groningen.

Zwart, C. Jan-Wouter. 2007. Some notes on the origin and distribution of the IPP-effect. *Groninger Arbeiten zur Germanistischen Linguistik* 45:77–99.

Zwart, C. Jan-Wouter. 2009. Relevance of Typology to Minimalist Inquiry. *Lingua* 119:1589–1606.

Zwart, C. Jan-Wouter. 2018. A note on the periphrastic past in afrikaans. *Stellenbosch Papers in Linguistics* 8:1–8.

Zwarts, Joost. 1992. X′ Syntax-x′semantics: on the interpretation of lexical and functional heads. Doctoral Dissertation, Utrecht Univeristy.

Samenvatting in het Nederlands

In deze dissertatie doe ik onderzoek naar semi-lexicaliteit in het verbale domain van het Nederlands en het Afrikaans. Een Nederlands voorbeeld van semi-lexicaliteit is gegeven in (1).

(1) Ik heb de hele dag ***zitten*** te **lezen**.

Zoals in het voorbeeld te zien is, is het houdingswerkwoord *zitten* in die zin gebruikt om progressief of duratief aspect aan te geven van het lexicale werkwoord van de zin (*lezen*). Het onderwerp van de zin is de hele dag aan het lezen geweest. Het feit dat het werkwoord *zitten* aspectuele informatie kan aangeven in een zin zoals die in (1), laat zien dat het werkwoord niet lexicaal maar functioneel gebruikt wordt in die zin. Echter, *zitten* is niet helemaal functioneel in zinnen zoals die in (1). Dit wordt duidelijk in een zin zoals die in (2).

(2) ??Ik heb de hele dag ***zitten*** te **zwemmen**.

De semantiek van het lexicale werkwoord van de zin, *zwemmen*, zijn incompatibel met een zittende positie. Een groot aantal sprekers van het Nederlands laten het niet toe om *zitten* te gebruiken om progressief of duratief aspect aan te duiden wanneer de semantiek van het lexicale werkwoord incompatibel is met de lexicale betekenis van *zitten* (Lemmens 2005, Haeseryn et al. 1997). Het feit dat veel sprekers van het Nederlands zinnen zoals (2) erg gek vinden, laat zien dat de lexicale semantiek van *zitten* nog steeds deels aanwezig is wanneer het gebruikt wordt om progressief of duratief aspect aan te duiden. Dit betekent dat in zinnen zoals (1) en (2), *zitten* zich als functioneel werkwoord gedraagt in de zin dat het het type aspect van het lexicale werkwoord aangeeft, terwijl het zich tegelijkertijd ook als lexicaal werkwoord gedraagt door het uiten van zijn eigen lexicale semantiek. Woorden die zowel functionele als lexicale eigenschappen hebben worden *semi-lexicale* woorden genoemd (Klockmann 2017). Het Nederlandse werkwoord *zitten* kan dus gezien worden als een woord dat semi-lexicaal gebruikt is. Het is echter belangrijk om op te merken dat *zitten* ook geheel lexicaal gebruikt kan worden (3).

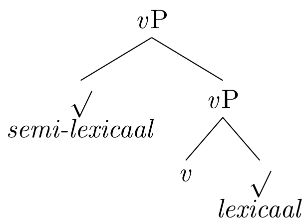
(3) Ik heb de hele dag op de bank ***gezeten***.

In deze zin is *zitten* het lexicale werkwoord van de zin: het onderwerp heeft de hele dag gezeten. In deze dissertatie maak ik een onderscheid tussen het *lexicale gebruik* – zoals in (3)– en het *semi-lexicale gebruik* – zoals in (1) – van een bepaald woord.

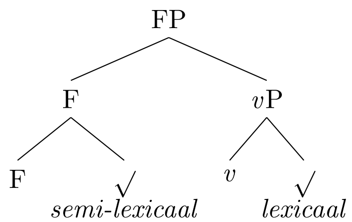
De theoretische bijdrage van deze dissertatie is als volgt. In dit werk presenteer ik een theoretische analyse van Nederlandse en Afrikaanse semi-lexicaal gebruikte werkwoorden. Daarbij gebruik ik het idee van frameworks zoals *Distributed Morphology* en*Exo-Skeletal Model* dat lexicale woorden de *spell out* zijn van *roots* (wortels), en functionele woorden de spell out van syntactische kenmerken op functionele hoofden (Halle en Marantz 1993, Harly en Noyer 1999, Borer 2005a). Zoals we hierboven hebben gezien, laat het semi-lexicale gebruik van een woord zowel functionele als lexicale kenmerken zien. De vraag is dus hoe semi-lexicaliteit geanalyseerd moet worden in termen van *roots* en syntactische kenmerken. In deze dissertatie houd ik het onderscheid aan tussen *roots* versus syntactische kenmerken voor lexicale en functionele woorden, maar stel voor dat onder bepaalde omstandigheden lexicale worden (*roots*) in het functionele domein van een ander lexicaal woord gepositioneerd kunnen zijn (cf. Klockmann (2017)). Wanneer dit gebeurt, wordt de *root* semi-lexicaal gebruikt.

Een centraal onderdeel van mijn theoretische voorstel is dat er twee fasen van semi-lexicaliteit bestaan. Beide fasen hebben verschillende onderliggende structuren. De eerste fase is bovendien de diachrone voorloper van de tweede fase. Dit betekent dat ik semi-lexicaliteit analyseer als het resultaat van grammaticalisatie (waarbij ik Haider (2001), Hagemeijer (2001), Klockmann (2017) en anderen volg). De twee fasen van semi-lexicaliteit moeten daarbij gezien worden als twee (zeer vroege) achtereenvolgende stappen in een grammaticalisatieprocess. In de eerste fase is de semi-lexicaal gebruikte *root* gepositioneerd in een zeer lage positie in het functionele domein van de lexicale *root*. In de tweede fase is de semi-lexicaal gebruikte *root* eerst samengenomen met een functioneel hoofd (F). Dit complexe hoofd is vervolgens in het functionele domein van de lexicale *root* geplaatst. De abstracte structuren van de twee fasen van semi-lexicaliteit (toegepast op het verbale domein – de focus van deze dissertatie) zijn weergegeven in (4) en (5).

(4) **Semi-lexical stage I**



(5) **Semi-lexical stage II**



De empirische focus van deze dissertatie is het verbale domein van het Nederlands en het Afrikaans. Specifiek kijk ik naar semi-lexicale *restructuring* (herstructurering) in deze twee talen. Grof gezegd is *restructuring* een fenomeen waarin twee werkwoorden dezelfde verbale projectie delen (cf. Wurmbrand (2001, 2004). Ik onderzoek in hoeverre semi-lexicale *restructuring* leidt tot morfosyntactische variatie en optionaliteit in allebei de talen.

Wat betreft het Nederlands onderzoek ik het semi-lexicaal gebruik van twee werkwoorden: *hoeven* en *zitten* (zoals in (1) hierboven). Wanneer deze werkwoorden semi-lexicaal gebruikt worden selecteren zij een infinitief die geïntroduceerd is door *te*, zoals geïllustreerd voor *hoeven* in (6).

(6) Hij zal morgen niet ***hoeven te*** **gaan** voetballen.

Gebaseerd op de resultaten van twee uitgebreide vragenlijsten laat ik zien dat er veel optionaliteit is tussen sprekers wat betreft de aanwezigheid en positie van *te*. Veel sprekers laten het bijvoorbeeld toe voor *te* om afwezig te blijven in de zin (7-a), of om op *hoeven* zelf geplaatst te zijn in plaats van op de infinitief die op *hoeven* volgt (waar het zou moeten staan) (7-b).

(7) a. Hij zal morgen niet ***hoeven*** **gaan** voetballen.
b. Hij zal morgen niet ***te hoeven*** **gaan** voetballen.

Voor *zitten* heb ik dezelfde feiten gevonden, zoals geïllustreerd in (8-a) en (8-b).

(8) a. Hij zal morgen lang op de bus moeten ***zitten*** **wachten**.
b. Hij zal morgen lang op de bus moeten ***te zitten*** **wachten**.

In deze dissertatie onderzoek ik de relatie tussen de mate van semi-lexicaliteit van deze twee werkwoorden (m.a.w., of ze in de eerste of tweede fase van semi-lexicaliteit zijn, en of ze in een transitie zijn van de eerste naar de tweede fase of niet) en de mate van morfosyntactische variatie en optionaliteit.

Ik doe hetzelfde voor constructies van *pseudocoordinatie* in het Afrikaans met het bewegingswerkwoord *loop* 'lopen' en de positiewerkwoorden *sit* 'zitten', *staan* 'staan' en *lê* 'liggen'. Het bewegingswerkwoord en de positiewerkwoorden zijn semi-lexicaal gebruikt in constructies van pseudocoordinatie, in de zin dat ze een type aspect van het lexicale werkwoord aangeven, maar tegelijkertijd hun eigen lexicale semantiek (deels) behouden. Gebaseerd op een corpusonderzoek en een vragenlijst laat ik zien dat er een hoge mate van morfosynactische variatie en optionaliteit bestaat in deze constructies. Twee voorbeelden hiervan zijn de aan- versus afwezigheid van de coordinator *en*, en de aan- en afwezigheid van de voltooid deelwoord markeerder *ge-*. Een voorbeeld met het bewegingswerkwoord *loop* is gegeven in (9).

(9) Ek het gister baie ***(ge-)loop* (*en*) praat**.
Ik heb gisteren veel GE-lopen en praten
'Ik heb gisteren veel (gelopen en) gepraat.'

Net zoals voor het Nederlands onderzoek ik de rol van de mate van semi-lexicaliteit van deze Afrikaanse werkwoorden in het bepalen van de hoeveelheid morfosyntactische variatie en optionaliteit in dit type constructies.

De opbouw van deze dissertatie is als volgt. Het theoretische voorstel wordt gepresenteerd in Hoofdstuk 2. In dit hoofdstuk bespreek ik twee recente benaderingen tot semi-lexicaliteit waarin het onderscheid tussen *roots* en syntactische kenmerken is geïmplementeerd, namelijk de benadering van De Belder (2011) en die van Klockmann (2017). In dit hoofdstuk introduceer ik ook het fenomeen *restructuring*, waarin ik specifiek Wurmbrand (2001)'s divisie tussen lexicale, semi-lexicale en functionele *restructuring* bespreek, die ik in deze dissertatie aanneem. Ik laat zien dat Nederlandse bewegings- en positiewerkwoorden (zoals semi-lexicaal *zitten* in (1)) een probleem vormen voor de analyse van semi-lexicale *restructuring* van Wurmbrand (2001). In de laatste sectie van dit hoofdstuk presenteer ik het hoofdonderdeel van mijn theoretische voorstel, namelijk dat we twee fasen van semi-lexicaliteit moeten onderscheiden, die verschillende onderliggende structuren hebben.

Het theoretische voorstel wordt vervolgens ingezet in twee casusstudies. De eerste casusstudie (Hoofdstukken 3–6) is geweid aan semi-lexicaliteit in het Nederlandse verbale domein. Specifieker gezegd onderzoek ik hier hoe de verschillende mates van semi-lexicaliteit van de Nederlandse werkwoorden *hoeven* en *zitten* een verschillende invloed hebben op de variatie en optionaliteit van de aanwezigheid en positie van de infinitiefmarkeerder *te* in werkwoordsclusters. De tweede casusstudie (Hoofdstukken 7–11) bekijkt semi-lexicaliteit in het Afrikaanse verbale domein. The focus van deze casusstudie is de impact van de mate van semi-lexicaliteit van bewegingswerkwoord *loop* 'lopen' en de positiewerkwoorden op de morfosyntactische variatie en optionaliteit in constructies van pseudocoordinatie. Drie vormen van variatie en optionaliteit worden onderzocht: the aan- versus afwezigheid van *en*, de aan- of afwezigheid van voltooid deelwoordmarkeerder *ge-*, en het type *Verb Second* dat de constructie kan ondergaan (*quirky* versus 'normaal' V2).

Ik sluit de dissertatie af met een algemene discussie in Hoofdstuk 12, waarin ik focus op de theoretische consequenties van mijn voorstel voor de theorie van grammaticalisatie binnen het Generatieve framework. Daarbovenop bespreek ik de verschillende types van morfosyntactische optionaliteit die waren gevonden in de twee casusstudies. Sommige types van optionaliteit zijn bijvoorbeeld het resultaat van grammaticalisatie die op het moment plaatsvindt, andere van bepaalde spell-out mechanismen, en andere van weer andere factoren. Ik eindig het hoofdstuk met een aantal mogelijke richtingen voor vervolgonderzoek.

CurriculumVitae

Cora Cavirani-Pots was born on 28 May 1990 in 's-Hertogenbosch, the Netherlands. In 2010, she started the bachelor program at the University of Utrecht, with two majors: modern literature and linguistics. Regarding the linguistics major, she specialised in language acquisition, with an internship and BA thesis at the Babylab (UiL-OTS). After that (2013-2015), she continued her studies in Utrecht at the UiL-OTS, with the research master linguistics. She specialised in syntactic variation, doing her internship and MA thesis with Prof. dr. Marjo van Koppen. In 2015 she applied for the PhD project titled 'Quality and Quantity in Linguistics' at KU Leuven, with which she started in January 2016 under supervision of Prof. dr. Jeroen van Craenenbroeck. In 2018, she spent two months as a visiting PhD student at the University of Cambridge, supervised by dr. Theresa Biberauer. This thesis is the result of her PhD research.