



Katholieke Universiteit Leuven
Faculteit Letteren
Departement Linguïstiek

Rood, groen, corpus!

*Een taalgebruiksgebaseerde analyse van
woordvolgordevariatie in tweeledige
werkwoordelijke eindgroepen*

Promotoren:
Prof. dr. D. Geeraerts
Prof. dr. D. Speelman

Proefschrift ingediend
tot het behalen van de
graad van Doctor in de
Taal- en Letterkunde:
Germaanse Talen door

Gert DE SUTTER

2005

Voor Katrien

Inhoud

Voorwoord	xiii
Overzicht statistische technieken	xv
1 Inleiding tot het onderzoek en het onderzoeksobject	1
1.1 Inleiding	1
1.2 Scope van het onderzoek	5
1.3 De groene en rode volgorde: een literatuuroverzicht	5
1.3.1 Algemene literatuur	5
1.3.1.1 Grammatica's	6
1.3.1.2 Taaladvies	9
1.3.1.3 Conclusie	12
1.3.2 Waarom de groene en rode volgorde naast elkaar bestaan: een overzicht van het diachroon-historische onderzoek	12
1.3.3 Taalwetenschappelijk onderzoek naar de implementatie van de woordvolgordevariatie in een formeel-autonoom grammaticamodel	18
1.3.3.1 Generatieve grammatica	18
1.3.3.2 Andere formele grammaticamodellen	24
1.3.3.3 Computatieve toepassingen	26
1.3.3.4 Conclusie	26

1.3.4	Taalgebruiksgeoriënteerd taalwetenschappelijk onderzoek	29
1.3.5	Taalwetenschappelijk onderzoek naar taalnormen	31
1.4	Doelstellingen en uitgangspunten van het onderzoek	34
1.5	Opbouw van de studie	38
2	Methodologische grondslagen van het onderzoek	45
2.1	Inleiding	45
2.2	Selectie van het materiaal: inhoud en structuur van de gebruikte corpora	46
2.3	Syntagmatische en paradigmatische restricties op het studieobject	48
2.4	Zoekopdracht	49
2.5	Zuivering van de dataset	50
2.6	Codering van de attestaties en statistische verwerking	53
3	De contextuele dimensie	59
3.1	Inleiding	59
3.2	De impact van regio	62
3.2.1	Status quaestionis	62
3.2.2	Hypotheses	67
3.2.3	Resultaten en discussie	67
3.2.3.1	Belgisch Nederlands versus Nederlands Nederlands	67
3.2.3.2	Streekverschillen	71
3.3	De impact van register	77
3.3.1	Status quaestionis	77
3.3.2	Hypotheses	79
3.3.3	Resultaten en discussie	80
3.3.3.1	Gesproken versus geschreven taal (modus)	81
3.3.3.2	Mate van interactie	83
3.3.3.3	Mate van redactionele controle	86
3.3.3.4	Modus, interactie en redactionele controle vergeleken en geherinterpreteerd	89
3.4	De interactie tussen het regio- en het registereffect	94

3.5	Samenvatting	96
3.5.1	Regio	97
3.5.2	Register	98
4	De prosodische dimensie	103
4.1	Inleiding	103
4.2	De impact van de accentverdeling in de werkwoordelijke eindgroep	108
4.2.1	Status quaestionis	109
4.2.2	Hypotheses	112
4.2.3	Resultaten en discussie	114
4.2.3.1	Scheidbaar samengesteld versus niet-scheidbaar samengesteld deelwoord	114
4.2.3.2	Afstand van het deelwoordaccent tot het begin van de eindgroep	117
4.2.3.3	Afstand van het deelwoordaccent tot het einde van de eindgroep	119
4.3	De impact van de accentverdeling in de linkercontext	121
4.3.1	Status quaestionis	121
4.3.2	Hypotheses	124
4.3.3	Resultaten en discussie	126
4.3.3.1	Definietheid van de laatste preverbale substantiefgroep	126
4.3.3.2	Informationaliteit van het laatste preverbale woord	129
4.3.3.3	Inherentie van de laatste preverbale constituent	131
4.3.3.3.1	Naamwoordelijk deel van het naamwoordelijk gezegde	132
4.3.3.3.2	Niet-werkwoordelijk deel van een werkwoordelijke vaste verbinding	134
4.3.3.4	Lengte van het middenstuk	153
4.3.3.5	Afstand van het deelwoordaccent tot het laatste preverbale woordaccent	155

4.3.3.6	Vergelijking van de verschillende operationaliseringen	158
4.4	De impact van de accentverdeling in de rechtercontext	160
4.4.1	Status quaestionis	161
4.4.2	Hypotheses	162
4.4.3	Resultaten en discussie	163
4.4.3.1	Aan- vs. afwezigheid van een laatste zinsplaats	163
4.4.3.2	Grammaticale relatie tussen de laatste zinsplaats en het hoofd	167
4.4.3.3	Afstand van het deelwoordaccent tot het eerste postverbale woordaccent	175
4.4.3.4	Vergelijking van de verschillende operationaliseringen	179
4.5	De interactie tussen de accentverdeling in, voor en na de werkwoordelijke eindgroep	184
4.6	Samenvatting	187
4.6.1	De accentdistributie in de werkwoordelijke eindgroep	188
4.6.2	De accentdistributie in de linkercontext	189
4.6.3	De accentdistributie in de rechtercontext	190
4.6.4	Slotopmerkingen	191
5	De semantische dimensie	197
5.1	Inleiding	197
5.2	De impact van de categoriale flexibiliteit van het deelwoord	198
5.2.1	De categoriale flexibiliteit van het deelwoord: algemene situering	199
5.2.2	Status quaestionis	200
5.2.3	Hypotheses	204
5.2.4	Operationalisering van de categoriale status van het deelwoord	205
5.2.4.1	Status quaestionis	205
5.2.4.2	Disambiguering van de categoriale status van het deelwoord	207

5.2.5	Resultaten en discussie	230
5.2.5.1	De categoriale status van het deelwoord (categorisch)	230
5.2.5.2	De categoriale status van het deelwoord (continu)	233
5.2.5.3	Tussentijdse samenvatting en interpretatie	239
5.2.5.4	De interactie tussen de participiale en constructionele semantiek	242
5.2.6	Aanvullingen bij de operationalisering van de categoriale status van het deelwoord	249
5.3	De impact van de aard van het hulpwerkwoord	251
5.3.1	Status quaestionis	251
5.3.2	Hypotheses	255
5.3.3	Resultaten en discussie	256
5.4	Samenvatting	259
5.4.1	Categoriale status van het deelwoord	260
5.4.2	Aard van het hulpwerkwoord	261
6	De psycholinguïstische dimensie	271
6.1	Inleiding	271
6.2	De impact van syntactische persistentie	272
6.2.1	Status quaestionis	273
6.2.2	Hypotheses	275
6.2.3	Resultaten en discussie	276
6.2.3.1	Voorafgaande tweeledige werkwoordelijke eindgroepen	277
6.2.3.2	Voorafgaande twee- en meerledige werkwoordelijke eindgroepen	279
6.2.3.3	De activatiegraad van voorafgaande werkwoordelijke eindgroepen	284
6.2.3.4	Onbedoelde syntactische priming in psycho- experimenteel onderzoek	287
6.3	De impact van de frequentie van het deelwoord	289

6.3.1	Status quaestionis	289
6.3.2	Hypotheses	290
6.3.3	Resultaten en discussie	290
6.4	Samenvatting	296
6.4.1	Syntactische persistentie	296
6.4.2	Frequentie van het deelwoord	298
7	Competitie en reïnterpretatie	301
7.1	Inleiding	301
7.2	Multivariate statistische analyse	303
7.2.1	Classificatieboomanalyse	303
7.2.2	Logistische regressie	309
7.3	Globale linguïstische analyse	318
7.3.1	Inhoudelijke round-up en reïnterpretatie	318
7.3.2	Naar een psycholinguïstische verklaring van de woordvolgordevariatie	321
7.3.2.1	Uitgangspunt	321
7.3.2.2	Een psycholinguïstische interpretatie van het frequentie-effect	323
7.3.2.3	Een psycholinguïstische interpretatie van het effect van de grammaticale relatie tussen de laatste zinsplaats en het hoofd	325
7.3.2.4	Slotbeschouwingen	327
8	Conclusies	333
	Appendices	339
	Literatuur	371

Voorwoord

Voor iedereen die op een directe of indirecte manier bijgedragen heeft tot de totstandkoming van dit doctoraatsonderzoek, zou ik op deze plaats een hartelijk dankjewel willen uitspreken.

In de eerste plaats gaat mijn dank uit naar mijn beide promotoren, Dirk Geeraerts en Dirk Speelman. Zij hebben mij gaandeweg wegwijs gemaakt in wat aanvankelijk nog de bijzonder chaotische en overweldigende wereld van de empirisch-kwantitatieve taalkunde leek. Voor de manier waarop ze dat gedaan hebben en de intensiteit waarmee, ben ik hen bijzonder dankbaar.

Ik dank ook Stef Grondelaers wiens onderzoek een lichtend voorbeeld was en wiens aanstekelijk enthousiasme mijn eigen onderzoek vaak in een stroomversnelling gebracht heeft. Mijn co-junior-QLVL'ers, Sofie Van Gijssel, Stefania Marzo, Ann Bertels, Koen Plevoets, Kris Heylen en José Tummers, ben ik dankbaar voor de warme collegialiteit en voor de vele gesprekken en discussies die ik met hen gehad heb, binnen en buiten het kader van de QLVL-vergaderingen. Heel wat ideeën die in dit proefschrift uitgewerkt zijn, zijn tot stand gekomen naar aanleiding van een van die gesprekken. Ik kan hen daarvoor niet genoeg bedanken.

Daarnaast bedank ik Sofie Van Gijssel, Karen Deschamps, Nele Nivelles, Kris Heylen, José Tummers en Geert Brône in het bijzonder, omdat ze zich bereid verklaard hebben om als informant te fungeren voor de 'disambigueringsenquête' (hoofdstuk 5).

Ik heb het geluk gehad om met een aantal collega's van gedachten te wisselen over aspecten of zelfs hele delen van mijn doctoraatsproject. Daarvoor wil ik de volgende personen van harte danken: A.Carbonez, G. Cornelissen, L. Draye, D. Drieghe, J. Goossens, W. Haeseryn, R. Hartsuiker, F. Joosten, H. Kloots, L. Kremer, K. Lahousse, A. Marynissen, P. Mertens, H. Smessaert, I. Van Canegem-Ardijns, R. van Hout en H. Van De Velde.

Ik ben ook het FWO-Vlaanderen erkentelijk, omdat het dit project met de nodige financiële middelen geschraagd heeft.

Voor de gezellige lunches, de ontspannende koffiekamergesprekken en de (meestal frustrerende) voetbalpartijtjes bedank ik Katleen, Hilde, Jan, Jan, Pieter, Michaël, Hans, Freek, Geert, Nele, Hendrik, Tine, Tom, Ana, Ilse, Ronny, Nele, Elke, Kristel, Pieter, An, Jean-Christophe. Lisette, Suzy, Marie-Louise, Ivo en Fred dank ik voor de technische en administratieve ondersteuning.

Dichter bij huis, voor de gezellige avonden en weekendjes uit: dankjewel An & Kristof, Stijn & Eveline, Bart & Veerle, Pieter, Filip & Kenny, Bart & Martien, Joris & Karolien, Kristel, Koen & Kaat, Sabine, Koen, Hanna & Simon, Rita, Koen & Anne, Anne Catherine & Wouter, Bert, Joke & Christophe, Bart & Annelies.

Heel speciaal dank ik mijn ouders en Magda en Mark voor de warme thuis in Oosterzele en Herent. Meer dan wie ook dank ik Katrien. Zij heeft niet alleen in de allerlaatste fase het hele proefschrift nog eens volledig nagelezen en de literatuurlijst geüpdatet, zij was er ook de hele tijd om mijn successen, frustraties en ontgoochelingen te delen. Als blijk van dank, voor alles en nog zoveel meer, draag ik het boek aan haar op.

Leuven, 20 april 2005

Overzicht statistische technieken

De volgende lijst geeft een overzicht van de statistische maten en technieken die in dit proefschrift aangewend zullen worden, samen met de gebruikte afkorting en de pagina waar de techniek voor het eerst geïntroduceerd en uitgelegd wordt.

Naam	Afkorting / Symbool	Uitleg
Breslow-Day-test	Br-D	p. 91
Chi-kwadraattest	χ^2	p. 68
Correlatiecoëfficiënt	r	p. 239
Classificatieboomanalyse	CBA	p. 303
Cohran-Mantel-Haenszeltest	CMH	p. 70
Gamma	γ	p. 88
Kappa	κ	p. 89
Lambda	λ	p. 88
Logistische-regressieanalyse	LRA	p. 309
Log-Likelihood ratio test	LLR	p. 136
Mantel-Haenszel chi-kwadraattest	MH- χ^2	p. 88
Odds ratio	o.r.	p. 69
Scheefheid	-	p. 147
Shapiro-Wilktest	-	p. 147
Wilcoxon rank sum test	-	p. 148

Hoofdstuk 1

Inleiding tot het onderzoek en het onderzoeksobject

1.1 Inleiding

In het verleden is al meermaals vastgesteld dat de schikking van de woorden in de zin, zelfs in talen met een relatief vaste woordvolgorde, aan aanzienlijke variatie onderhevig kan zijn. Taalwetenschappers hebben zich daarom vanuit verschillende achtergronden, met verschillende doelstellingen en verschillende methodes gebogen over allerlei fenomenen van syntactische variatie. Daarbij kan grofweg een onderscheid gemaakt worden tussen onderzoeken die zich concentreren op wat we *interconstituentenvariatie* zouden willen noemen en onderzoeken die zich bezighouden met *intraconstituentenvariatie*. Het eerste type onderzoek heeft de alternatie tussen twee of meer constituenten als onderwerp, het tweede type de alternatie tussen twee of meer elementen binnen een constituent. Het hoeft nauwelijks gezegd te worden dat interconstituentenvariatie vaker voorkomt en daarom ook vaker het onderwerp van onderzoek geweest is, denken we maar aan de datiefalternatie, de relatieve volgorde van prepositionele constituenten en/of nominale constituenten, Heavy-NP-shift en andere extrapositiefenomenen (o.a. Givón 1988, Dik 1989, Siewierska 1993, Hawkins 1994 & 2000, Wasow 2002). Het onderzoek naar die fenomenen van interconstituentenvariatie is al vanuit diverse perspectieven bekeken—sociologisch, stilistisch, geografisch, cognitief, fonetico-fonologisch, prosodisch, (morfo-)syntactisch, semantisch en/of discursief—en heeft een aantal belangrijke inzichten opgeleverd over de onderliggende cognitieve processen bij zinsverwerking

en de functie van syntactische optionaliteit, zoals ambiguïteitsvermijding, markering van de pragmatische waarde en het faciliteren van de cognitieve verwerking (o.a. Hawkins 1994, Gibson 1998, Arnold et al. 2000, Clifton & Duffy 2001, Wasow 2002). Het onderzoek naar de intraconstituentenvariatie, dat van een fundamenteel andere aard is, is daarentegen minder prominent aanwezig en de variatie lijkt ook minder makkelijk te beregelen. Illustratief daarvoor is dat de verklaringsmechanismen die voor fenomenen van interconstituentenvariatie voorgesteld zijn, slechts in beperkte mate toegepast kunnen worden op intraconstituentenvariatie en de kwantitatieve woordvolgordematen zoals die ontwikkeld zijn door Givón (1988), Hawkins (1994) en Wasow (2002) niet gebruikt kunnen worden om de woordvolgorde binnen een constituent te voorspellen.

De voorliggende studie is gewijd aan een dergelijk type syntactische variatie dat bekend is voor zijn schijnbaar onontwarbaar kluwen van interagerende krachten, met name de woordvolgordevariatie in wat een van de best bestudeerde syntactische constructies van de Nederlandse taal moet zijn: de werkwoordelijke eindgroep.¹ Daarmee wordt traditioneel het groepje werkwoorden bedoeld dat achteraan in de zin de tweede pool van de tangconstructie bezet, met name de infinitie werkwoorden in een hoofdzin en de finiete en infinitie werkwoorden in een bijzin (ANS 1997: 1225-1235; zie voorbeeldzin (1a) en (1b)).² De eindgroepen staan er bekend om dat de constituerende delen ervan in meer of mindere mate van plaats kunnen wisselen, afhankelijk van de aard van de onderdelen en de lengte van de eindgroep. Vanuit verschillende invalshoeken en met de meest uiteenlopende methodes zijn verschillende types werkwoordelijke eindgroepen bestudeerd: twee-, drie- of meerledige eindgroepen, eindgroepen met alleen infinitieven of deelwoorden, eindgroepen met zowel infinitieven als deelwoorden en eindgroepen die doorbroken worden door een niet-werkwoordelijk element. Een bijzonder prominente plaats binnen de onderzoekstraditie naar de volgordevariatie in de werkwoordelijke eindgroep werd daarbij toebedeeld aan de tweeledige werkwoordelijke eindgroep die opgebouwd is uit een hulpwerkwoord [AUX] en een deelwoord [PART]:³

(1a) [...] dat moordende chauffeurs van de weg gehaald worden. (HLN)⁴

(1b) [...] dat moordende chauffeurs van de weg worden gehaald. (HLN)

De syntactische variant in de eerste voorbeeldzin wordt traditioneel de *groene volgorde* [PART+AUX] genoemd, die in de tweede zin de *rode volgorde* [AUX+PART].⁵ Dat type woordvolgordevariatie heeft de afgelopen vijftig jaar in de neerlandistiek heel wat pennen in beweging gebracht. Over het algemeen worden de studies van Van den Berg (1949) en Meeussen & Vanacker (1951), maar vooral de dialectologische studie van Pauwels (1953), waarin de termen *groene* [PART+AUX] en *rode* [AUX+PART] *woordvolgorde* geïntroduceerd werden, als startpunt beschouwd van de taalwetenschappelijke bestudering van de syntactische flexibiliteit in de tweeledige werkwoordelijke eindgroep. In de daaropvolgende decennia zouden tal van wetenschappers van verschillend slag verschillende aspecten van die woordvolgordevariatie bestuderen, hetgeen tot een niet onaanzienlijke stroom van publicaties heeft geleid. Dat het beginpunt van de taalwetenschappelijke studie van de volgordevariatie omstreeks 1950 gesitueerd wordt, betekent geenszins dat er voor die tijd geen aandacht besteed werd aan de volgorde van de werkwoorden in de werkwoordelijke eindgroep. De aandacht bleef echter meestal beperkt tot de vermelding van de loutere mogelijkheid tot alternantie, hier en daar vergezeld van een al dan niet door impressionistische argumenten vergezeldde voorkeur voor een van de woordvolgordes (zie bv. Bilderdijk, Mulders, Brill en anderen in Van der Horst 2000 en Van der Horst 2002: 547 e.v.). In de eerste helft van de twintigste eeuw kwam daar geleidelijk aan verandering in wanneer linguïsten aandacht begonnen te krijgen voor regionale en ritmische correlaten van de variatie (bv. Overdiep 1937, 1949). Die ontwikkeling mondde uiteindelijk uit in de studie van Pauwels (1953), die als eerste een volledig onderzoek wijdde aan de volgordevariatie in de werkwoordelijke eindgroep.

De voorliggende studie ambieert in het algemeen de onderliggende mechanismen op te sporen die de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde in het hedendaagse Nederlands beïnvloeden. Dat we daarbij niet in het luchtledige starten, mag al gebleken zijn uit de beknopte historische schets hierboven. Vooraleer we met het eigenlijke onderzoek van start gaan, zal in de rest van dit hoofdstuk dan ook eerst een overzicht gegeven worden van de algemene en wetenschappelijke literatuur van de voorbije decennia (paragraaf 1.3). In een eerste luik zullen we daarbij kijken of en in welke mate grammatica's en taaladviesliteratuur de volgordevariatie in de tweeledige werkwoordelijke eindgroep bespreken. Op die

manier kunnen we relatief snel te weten komen hoe de taalkundige en talige gemeenschap tegen de variatie aankijkt, of het een hot topic is, of een van de varianten afgekeurd wordt, of een van de varianten de voorkeur krijgt en in hoeverre de taalwetenschappelijke resultaten, die we in de paragrafen daarna zullen bespreken, doorgestroomd zijn naar de algemene literatuur. In een tweede luik gaan we wat dieper in op de resultaten van het wetenschappelijke onderzoek naar de syntactische variatie. In een eerste stap zullen we aan de hand van het diachroon-historische onderzoek het ontstaan en de ontwikkeling van de beide varianten schetsen. Niettegenstaande dat deze studie uitsluitend een synchronon perspectief op de variatie heeft, kan de kennis van de varianten in vroegere taalfasen erg belangrijk zijn bij de interpretatie van de synchrone resultaten, bijvoorbeeld als er sprake zou zijn van een aan de gang zijnde taalverandering (vgl. bv. Haeseryn 1990: 42). In een tweede, synchronon-georiënteerde, stap gaan we dieper in op de onderzoeksinspanningen die geleverd zijn binnen het paradigma van de op het taalvermogen (*competence*) gerichte, autonomistische, formele taalkunde, die zich vanaf de jaren '70 intensief met woordvolgordevariatie beziggehouden heeft. Daarbij wordt ruim aandacht besteed aan de belangrijkste generatief-taalkundige analyses, maar er wordt ook plaats ingeruimd voor verwante formalistische benaderingen, zoals bijvoorbeeld *Head-driven Phrase Structure Grammar* en *Lexical-Functional Grammar*. Niettegenstaande dat onze studie op een aantal punten wezenlijk verschilt van die formele analyses, is een bespreking van de wetenschappelijke literatuur in dat domein om twee redenen interessant: enerzijds trachten we daarmee te voorzien in een representatief beeld van de verscheidenheid aan analyses die in de loop van de afgelopen decennia uitgevoerd zijn, maar anderzijds, en voornamelijk, trachten we aan de hand van dat overzicht aan te geven waar onze basisuitgangspunten en basisdoelstellingen verschillen van die van de formele taalkunde. In een laatste, eveneens synchronon-georiënteerde, stap gaan we wat dieper in op de resultaten van de verschillende empirische onderzoeken (naar taalgedrag in 1.3.4 en taalnormen in 1.3.5). Daarbij zal een algemene karakterisering van het onderzoeksveld gegeven worden en zal kort ingezoomd worden op de verklaringsdimensies die men kan onderscheiden in het onderzoek naar de volgordevariatie in tweeledige werkwoordelijke eindgroepen. Het huidige hoofdstuk wordt afgesloten met een gedetailleerde beschrijving van de uitgangspunten en doelstellingen (paragraaf 1.4) en de opbouw van de dissertatie (paragraaf 1.5).

Vooraleer we van start gaan met het overzicht van de algemene en wetenschappelijke literatuur, staan we eerst nog even stil bij de scope van dit onderzoek.

1.2 Scope van het onderzoek

Zoals in de inleiding al aangegeven werd, wil deze studie een antwoord bieden op de vraag welke mechanismen er in het hedendaagse Nederlands voor zorgen dat taalgebruikers nu eens de groene [PART+AUX] en dan weer de rode [AUX+PART] volgorde kiezen. Aan de hand van reëel geattesteerd taalmateriaal uit twee verschillende corpora zullen we daartoe het effect van een set van taalinterne en taalexterne factoren op de keuze voor een van de woordvolgordes empirisch en kwantitatief bestuderen. Hoewel de meeste factoren die in deze studie besproken zullen worden ook al in vroegere studies aan bod kwamen, verschilt de huidige studie toch grondig van de vorige studies doordat ze al die factoren in een enkele studie samenbrengt, elke factor afzonderlijk onderwerpt aan een diepteanalyse, de verhouding tussen de verschillende factoren onderling bestudeert en het simultane effect van alle factoren op de woordvolgorde nagaat. Veel meer dan in vroegere studies het geval was, wordt daarbij de nadruk gelegd op de methodologische en analytische vereisten voor gedegen wetenschappelijk onderzoek, als noodzakelijke voorwaarde voor inhoudelijke progressie. Die inhoudelijke progressie zal er enerzijds in bestaan dat vroeger gepresenteerde verklaringsprincipes op hun soliditeit getest worden en met elkaar vergeleken worden en dat anderzijds een nieuw verklaringsprincipe in psycholinguïstische termen geïntroduceerd wordt. De doelstellingen, uitgangspunten en de opbouw van deze studie worden in meer detail toegelicht in paragraaf 1.4.

1.3 De groene en rode volgorde: een literatuuroverzicht

1.3.1 Algemene literatuur

Vanuit de idee dat de algemene literatuur als graadmeter voor de verankering van een bepaald fenomeen in de taalkundige en talige gemeenschap kan fungeren, geven we in deze paragraaf een overzicht van de belangrijkste tendensen, overeenkomsten en verschilpunten die opgemerkt kunnen worden in de niet specifiek op taalkundigen

gerichte algemene literatuur. Het onderstaande overzicht wil echter op geen enkele wijze aanspraak maken op exhaustiviteit, maar het heeft wel als doel om een representatief beeld te geven van de positie van de volgordevariatie in de algemene literatuur van de 20^{ste} eeuw en het begin van de 21^{ste} eeuw. Daarbij wordt een onderscheid gemaakt tussen *grammatica's* en wat we met een algemene benaming *taaladviesliteratuur* zouden willen noemen (stijlboeken, taalverzorgings- en taalzuiveringswerken en dies meer).⁶ De onderzochte literatuur hebben we ten behoeve van een snelle raadpleging opgenomen in een aparte literatuurlijst (appendix A). Werken die de hier bestudeerde variatie vermelden, hebben we in appendix A gemarkeerd met een asterisk.

1.3.1.1 *Grammatica's*

De grammatica die het uitvoerigst op de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep ingaat, is de *Algemene Nederlandse Spraakkunst*, waarvan de eerste editie in 1984, en een volledig herwerkte, tweede editie in 1997 verscheen. Daar waar de eerste editie (ANS 1984: 599) zich beperkt tot de vermelding dat de beide varianten evenwaardig zijn en dat de keuze bepaald wordt door regionale en stilistische factoren, gaat de tweede editie van de ANS (1997: 1067-1069) dieper in op andere determinerende factoren, evenwel zonder de precieze aard van de invloed van die factoren te bespreken, geheel conform de doelstellingen die de ANS (1997: 5-6) zichzelf stelt. Zo wordt gezegd dat de keuze voor een van de volgordes in zekere mate individueel bepaald is en verder beïnvloed wordt door geografische, stilistische en zelfs ritmische factoren. De stilistische factor uit zich in het feit dat in gesproken taal vaker de groene volgorde [PART+AUX] gebruikt wordt en in geschreven taal, en dat geldt zeker voor journalistieke teksten, de rode volgorde [AUX+PART]. De ritmische factor (in andere grammatica's ook wel melodie of intonatie genoemd) wordt in verband gebracht met een niet exact te beschrijven evenwichtige klemtoonverdeling aan het eind van de zin, waarbij de accentuering aan het einde van de zin, het accentpatroon dat scheidbaar van niet-scheidbaar samengestelde werkwoorden⁷ (i.c. deelwoorden) onderscheidt en de beklemtoning van de constituent die voorafgaat aan de eindgroep een rol lijken te spelen. De ANS merkt daarbij ook nog op dat bij de behandeling van de woordvolgordeflexibiliteit een onderscheid gemaakt moet worden tussen werkwoordelijke en naamwoordelijke gezegdes waarvan het deelwoord tot

adjectief geworden is. In het eerste type gezegde zijn nl. steeds de beide volgorde mogelijkheden mogelijk, terwijl het tweede type in principe enkel in de groene volgorde [PART+AUX] kan voorkomen. Toch, zo stelt de ANS vast, gebeurt het (zij het onterecht) dat door analogiewerking de rode volgorde [AUX+PART] in naamwoordelijke gezegdes gebruikt wordt.

Andere invloedrijke grammatica's⁸ zoals Paardekooper en Paardekooper et al. (1960, 1963, 1971, 1975, 1977) en Rijpma & Schuringa (1978) vermelden louter dat het deelwoord en het hulpwerkwoord van plaats kunnen wisselen, Overdiep (1937, 1949), Rijpma & Schuringa (1957), Van der Lubbe (1965), Den Hertog (1973) en De Schutter & Van Hauwermeiren (1983) gaan daarbij iets verder door een of meerdere van de hierboven vermelde factoren aan te halen of door te verwijzen naar resultaten van wetenschappelijk onderzoek. Van wat als invloedrijke grammatica beschouwd kan worden, zeggen Overdiep (1928), De Groot (1949) en Van den Toorn (1978, 1982) niets over de woordvolgordevariatie in eindgroepen met een deelwoord en een hulpwerkwoord.

Als ook de andere geconsulteerde grammatica's in rekening gebracht worden (57 in het totaal; zie appendix A), dan zien we dat iets meer dan de helft van alle geconsulteerde grammatica's (33/57) vermeldt dat er variatie is; de 24 andere grammatica's gaan niet in op de volgordekwestie. Indien we de grammatica's ordenen volgens een temporele dimensie, dan valt op dat van de 13 grammatica's uit de periode 1861-1945 er 7 de volgordevariatie vermelden; in de periode 1945-1970 loopt dat aantal op tot maar liefst 13 van de 16 grammatica's, waarna het opnieuw daalt in de periode na 1970 naar 15 van de 28. Of de stijging van de aandacht in de periode 1945-1970 representatief is voor alle grammatica's, kan niet uitgemaakt worden zonder een diepgaandere systematische studie, maar opvallend is wel dat die piekperiode samenvalt met de eerste wetenschappelijke analyses van de volgordevariatie. Of dat verband toevallig is of niet, of we een causaal verband tussen die twee feiten moeten leggen en in welke richting dat causaal verband moet gaan (heeft de grammaticabeschrijving de taalwetenschappelijke studie gestimuleerd of net omgekeerd?), moet nog onderzocht worden. Een ander opmerkelijk feit is dat van de 16 grammatica's die een Belgische uitgever hebben (de grammatica's die zowel een Nederlandse als een Belgische uitgever hebben, zijn hier als 'Belgisch' beschouwd) er

maar liefst 14 iets zeggen over de volgordekwestie (versus 19/41 voor de Nederlandse grammatica's).⁹ Het ligt voor de hand om die opvallende tegenstelling terug te voeren op de ruime aandacht die in Vlaanderen besteed wordt aan de werkwoordelijke eindgroep, en meer bepaald aan de doorbreking van de werkwoordelijke eindgroep (type *dat hij zich heeft schuldig gemaakt*) en de tussenplaatsing van een deelwoord in een drieledige werkwoordelijke eindgroep (type *dat hij dat kan gedaan hebben*). Die fenomenen zijn in Vlaanderen steviger verankerd dan in Nederland (zie bv. Haeseryn 1990), hetgeen ook tot uiting komt in het feit dat de norm in Nederland pas later opduikt (Van der Horst 1999: 540). Daardoor is het niet ondenkbaar dat de drang naar informatie over allerhande volgordekwesties in de werkwoordelijke eindgroep gestimuleerd werd.

Opvallend is verder nog dat geen van de onderzochte grammatica's een van de volgordevarianten afkeurt, al zegt Paardekooper wel expliciet dat zijn voorkeur naar de groene woordvolgorde [PART+AUX] uitgaat, noemt Den Hertog de groene volgorde [PART+AUX] vloeiender en vindt Van den Berg (1967), impliciet ook al in Van den Berg (1957), dat de groene volgorde [PART+AUX] het meeste aansluit bij het grondbeginsel waarbij het hulpwerkwoord als afsluitend element van de bijzin beschouwd wordt.

Naast grammatica's die in eerste instantie op moedertaalsprekers gericht zijn, werden ook 16 grammatica's voor anderstaligen en 3 grammaticale lexica bekeken. 6 grammatica's voor anderstaligen maakten melding van de variatie in de werkwoordelijke eindgroep. Behalve in Brachin (1977) en Brachin & Vincent (1985) wordt daarbij nergens ingegaan op eventuele taalinterne of taalexterne verschillen tussen de groene [PART+AUX] en rode [AUX+PART] woordvolgorde. Brachin & Vincent zelf vermelden kort een potentiële stilistische en regionale factor, maar besluiten ten slotte toch dat "usage is far from consistent, so that one fails to see why in a particular case, the author has preferred the 'green' construction [...] and in another the 'red' [...]. What one in fact finds, then, is completely free variation" (Brachin & Vincent 1985: 56-58). Van de drie onderzochte lexica, vermeldt enkel die van Pensaert (1991) de volgordevariatie en geeft daarbij mee dat de groene volgorde [PART+AUX] spreektaalig is en de rode [AUX+PART] schrijftalig.

1.3.1.2 Taaladvies

Alhoewel er zeker verschillen zijn tussen (i) de stijlboeken die verschillende kranten- en televisieredacties, zoals *NRC-Handelsblad*, *Trouw* en *De Volkskrant* in Nederland en *De Standaard* en de *VRT* in Vlaanderen uitgegeven hebben, en (ii) een werk als dat van Van de Pol (1994), *In goed Nederlands: overzichtelijke taalgids met actuele voorbeelden uit de krant; voor op het werk, op school en thuis*, zowel inhoudelijk als wat de doelstellingen betreft, behandelen we de werken die allerhande taaladvies verstrekken samen. De aard van dat advies kan verschillende vormen aannemen, afhankelijk van het doel dat de auteur zich gesteld heeft en het doelpubliek dat hij beoogt, gaande van de meest rabiate vorm van taalzuivering over taalverzorging naar een objectiever ogend taaladvies. Aangezien het moeilijk is om voor elk werk de precieze doelstelling na te gaan en aangezien de verschillen tussen de net genoemde types van advies veeleer gradueel van aard zijn, en daarom vage grenzen hebben, behandelen we deze types van taaladvisering samen. Het spreekt vanzelf dat de hierboven besproken grammatica's ook een zekere normatieve, taaladviserende functie hebben (zie bv. ANS 1997: 5-6), maar ze verschillen van de hier behandelde taaladviserende werken, doordat ze een min of meer volledige en systematische behandeling geven van de grammatica en zich niet bij voorbaat richten op zogenaamd smeulende taalkwesties.

In totaal werden 89 taaladviserende werken geraadpleegd. Ook hier geldt dat we met die 89 werken op geen enkele manier aanspraak maken op exhaustiviteit¹⁰. Op basis van het literatuuroverzicht van Haeseryn (1999) blijkt echter wel dat de belangrijkste taaladviserende werken uit Vlaanderen en Nederland in ons overzicht opgenomen zijn. Circa een derde van die werken vermeldt het bestaan van de volgordevariatie (32/89). Dat is weliswaar minder dan de vermeldingsgraad in de grammatica's, maar is al bij al bijzonder veel voor een grammaticale kwestie. Eerder was nl. al vastgesteld dat de aandacht van de meeste taaladviseurs gericht was op lexicale, idiomatische en uitspraakkwesities, zeker tijdens de hoogdagen van het Vlaamse integrationistische taalzuiveringsbeleid (Taeldeman 1992: 39, Haeseryn 1996: 111-112). Een van de uitzonderingen op die regel was en blijft de volgordevariatie in werkwoordelijke eindgroepen, die blijkens allerhande vragen die binnenkomen bij allerlei taaladviserende instanties wel tot het taalbewustzijn van de

taalgebruiker is doorgedrongen (zie Haeseryn 1990: 23-24), misschien precies omdat er een behoorlijk hoge graad van uniformiteit op syntactisch vlak lijkt te zijn (Haeseryn 1990: 15, Taeldeman 1992: 47). Als we de taaladviserende werken verdelen over de periodes die we daarnet introduceerden, dan kan een geleidelijke stijging van de aandacht voor de volgordevariatie in tweeledige werkwoordelijke eindgroepen met een deelwoord en een hulpwerkwoord vastgesteld worden: van 1 van de 9 in de periode voor 1945, over 3 van de 13 in de periode 1945-1970 naar 27 van de 64 in de periode na 1970.¹¹ Hieruit blijkt dat de taaladviesverstrekking over de woordvolgordekwestie pas goed van start gaat vanaf de jaren '70, na twee decennia taalwetenschappelijk onderzoek, in de bloeiperiode van de taaladviesverlening. Als de werken opgesplitst worden volgens regio van de uitgever¹² dan is het duidelijk dat in Vlaanderen meer aandacht besteed wordt aan de kwestie (13/35 versus 18/53 voor Nederland). De reden voor deze ruimere aandacht kan, zoals hierboven reeds gesuggereerd werd, in verband gebracht worden met de aandacht voor een voornamelijk Belgisch-Nederlandse karakteristiek, m.n. de doorbreking van de werkwoordelijke eindgroep en de plaatsing van het deelwoord tussen twee hulpwerkwoorden in een drie- of meerledige eindgroep.

De aard van het advies over de groene [PART+AUX] en rode [AUX+PART] volgorde varieert nogal sterk. Naast de werken die enkel zeggen dat de keuze volledig vrij is (bv. D'Haenens et al. 1991), kunnen er nog twee andere types onderscheiden worden: een eerste type adviseert een bepaalde volgorde, zonder de andere volgorde bij voorbaat uit te sluiten; een tweede type weerlegt een aantal vooroordelen en/of schetst een aantal factoren die de keuze voor de een of andere volgorde kan beïnvloeden. Tot het eerste type behoren Charivarius¹³ (s.d. & 1945) en het stijlboek van *de Volkskrant* (Van Gessel & Hulsbosch 1992) voor wie de rode volgorde [AUX+PART] beter is dan de groene [PART+AUX], Taalbaak adviseert bij twijfel de rode volgorde [AUX+PART], terwijl het stijlboek van de *VRT* (Hendrickx 2003) de keuze voor de groene volgorde [PART+AUX] het veiligste acht.¹⁴ Dat de groene volgorde [PART+AUX] veiliger is, heeft wellicht alles te maken met uiteenlopende grammaticaliteitsoordelen in het geval van een rode cluster [AUX+PART] waarbij het deelwoord adjectivische eigenschappen heeft. In dat geval zijn alle taaladviserende werken het erover eens dat de rode volgorde [AUX+PART] verboden is (bv. Overduin 1986, Tiggeler 2000). Tot het tweede type behoren onder meer die werken die de

traditionele mythe ontcrachten dat de groene volgorde [PART+AUX] vermeden moet worden, omdat het een germanisme zou zijn (Hermkens 1966, De Boer 1972, Klein 1992). In de Duitse Hochsprache is namelijk enkel de groene volgorde [PART+AUX] mogelijk en, gegeven het feit dat het Nederlands ook over de rode volgorde [AUX+PART] beschikt, heeft dat een aantal mensen ertoe aangezet om de groene volgorde [PART+AUX] als germanisme te beschouwen. Wie dat voor het eerst gesuggereerd heeft, is niet helemaal duidelijk. Stroop (1970: 134) merkt wel op dat het niet later dan 1960 geweest kan zijn, getuige het volgende citaat: de “‘Gids voor de redactie’ van het Algemeen Handelsblad [meedeelt] dat de volgorde: ‘overschreden hebben’ en ‘beschouwd worden’ Duits is en gemeden dient te worden” (zie ook bij Haeseryn 1990: 40-41 en De Schutter 1996: 208, voetnoot 2 voor andere literatuur). Een aantal andere werken van het tweede type poneert verder nog dat de groene volgorde [PART+AUX] de oudste is (Overduin 1986; zie 1.3.2), dat de groene volgorde [PART+AUX] globaal genomen frequenter is (Heidbuchel 1967, 1975, De Boer 1972) en dat er een regionaal verschil is, in die zin dat Vlaanderen vaker dan Nederland kiest voor de groene volgorde [PART+AUX] (Buyse & Penninckx 1977, 1984). Op microregionaal vlak worden er tegenstrijdige uitspraken gedaan: zo beweren Renkema (1995) en Lamers (1997) dat in het oosten het vaakst de groene volgorde [PART+AUX] gebezigd wordt, terwijl van Sterkenburg & Van den Toorn 1997 net het omgekeerde beweren (zie 3.2.1 voor de wetenschappelijke bevindingen). Tot slot worden ook nog registerverschillen aangehaald, m.n. in de gesproken taal is er een voorkeur voor de groene volgorde [PART+AUX], in de geschreven taal voor de rode volgorde [AUX+PART]—Van Sterkenburg & Van den Toorn 1997 menen dat de rode volgorde [AUX+PART] prestigieuzer is (zie ook Michels 1957: 111), waardoor het vaak in de journaaltaal gebruikt wordt (Klein & Visscher 1992 & 1996 spreken zelfs over een angst voor de groene volgorde [PART+AUX]). Buyse & Penninckx (1984) adviseren desondanks de rode volgorde [AUX+PART] te gebruiken als men niet-plechtig schrijft, anderen zeggen dat de keuze bepaald moet worden door het gevoel voor ritme en melodie (o.a. Charivarius s.d. & 1945, Heidbuchel 1967 & 1975, Van Eijk 1984, Buyse & Penninckx 1984, Lamers 1997) of met het oog op voldoende stilistische afwisseling (Heidbuchel 1967, Buyse & Penninckx 1977 & 1984; zie echter Michels 1957: 107-108). In de volgende hoofdstukken zullen we de meeste van de beweringen die in het tweede type taaladviesliteratuur geuit zijn aan de hand van

het beschikbare taalwetenschappelijk onderzoek en de resultaten van onze eigen analyses toetsen.

1.3.1.3 Conclusie

Samenvattend kan gezegd worden dat de volgordevariatie in de tweeledige werkwoordelijke eindgroep met een deelwoord en een hulpwerkwoord ongeveer in de helft van de geraadpleegde grammatica's en in een derde van de taaladviserende literatuur vermeld wordt. Geen van de beide varianten wordt echt afgekeurd, al wordt sporadisch wel een voorkeur uitgesproken voor de groene [PART+AUX] of de rode [AUX+PART] volgorde. Opvallend is wel dat de taalwetenschappelijke resultaten in min of meerdere mate hun weg naar de algemene literatuur gevonden hebben—al kan niet uitgesloten worden dat de auteurs van de algemene werken ook hun eigen intuïties aangesproken hebben: er wordt nl. gewag gemaakt van gebruiksverschillen die afhankelijk zijn van het type gezegde, het ritme, de regio, het register en dergelijke meer. Men staat daarbij echter niet altijd op dezelfde lijn wat de precieze invloed is van een factor (zie bv. regio en register).

In de volgende paragrafen (1.3.2 t.e.m. 1.3.5) maken we de sprong van de algemene naar de wetenschappelijke literatuur. In paragraaf 1.3.2 komt eerst een overzicht aan bod van het onderzoek naar de historische wortels en de ontwikkeling van de syntactische variatie. In de paragrafen die daarop volgen, verschuift de focus van het historisch-diachrone naar het zuiver synchron-hedendaagse: zoals aangekondigd worden achtereenvolgens de verwezenlijkingen van de formele taalkunde, de taalgebruiksgeoriënteerde taalkunde en de taalnormen- en taalattitudesgeoriënteerde taalkunde voorgesteld.

1.3.2 Waarom de groene en rode volgorde naast elkaar bestaan: een overzicht van het diachroon-historische onderzoek

De vraag naar de historische ontwikkeling van de beide varianten is zo oud als de eerste wetenschappelijke analyses van het fenomeen (Van den Berg 1949). Dat ligt ook voor de hand, niet alleen omdat taalvariatie onmiddellijk de gedachte aan taalverandering oproept, maar vooral omdat taalwetenschappers getracht hebben om vanuit vroegere taalfasen van het Nederlands een antwoord te formuleren op de vraag

waarom de beide syntactische alternatieven, waarvan wordt aangenomen dat er nauwelijks een grammaticaliteits- of betekenisverschil is, in complexe verhoudingen naast elkaar bestaan. Uit het beperkte historisch-diachrone onderzoek blijkt alvast duidelijk dat de twee volgordes al eeuwenlang, in wisselende verhoudingen, naast elkaar bestaan, meer bepaald al vanaf de dertiende eeuw—en misschien zelfs al vroeger—zoals onder andere blijkt uit de steekproeven van een aantal ambtelijke teksten uit de dertiende tot de zeventiende eeuw (Van den Berg 1949).¹⁵ Van den Berg heeft daarbij niet de intentie om de beide volgordevarianten middels frequentietellingen tegen elkaar af te wegen, laat staan beïnvloedende factoren te identificeren, maar wel om een mogelijke ontwikkeling van de beide volgordevarianten doorheen de eeuwen te schetsen. Van den Berg (1949) acht daarin twee concurrerende tendensen belangrijk, m.n. de tendens tot achteropplaatsing van het finiete werkwoord in de bijzin en de ontwikkeling van een perifrastische werkwoordstijd, waarbij de betekenis van het deelwoord verandert van een predicatieve bepaling naar een onderdeel van een complexere werkwoordstijd.¹⁶ Het ultieme resultaat van de eerste ontwikkeling is het finiete werkwoord helemaal achteraan in de zin (resultierend in de groene volgorde [PART+AUX]), het resultaat van de tweede ontwikkeling is het finiete werkwoord op de voorlaatste plaats (resultierend in de rode volgorde [AUX+PART]). Deze concurrerende tendensen zouden er volgens Van den Berg (1949) toe geleid hebben dat de groene volgorde [PART+AUX] de enig mogelijke volgorde zou worden, gegeven de tendens tot achteropplaatsing van het finiete werkwoord en gegeven het feit dat veel deelwoorden als adjectief gebruikt worden, waardoor ze net als alle andere bepalingen voor het finiete werkwoord geplaatst moeten worden. Dat de rode volgorde [AUX+PART] in de loop der tijden toch is blijven bestaan, heeft volgens Van den Berg (1949) te maken met de conformiteit aan de schrijftaaltraditie. Aanwijzingen voor die hypothese vindt hij in zijn dialectologische onderzoekingen op basis van het materiaal in *Winkler's Dialekticon* en *Van de Schelde tot de Weichsel* (van J. & L. Leopold, zie Van den Berg 1949: 155). Daaruit blijkt dat enkel de gebieden met een onafgebroken schrijftaaltraditie de beide volgordes zijn blijven gebruiken.

Duinhoven (1998) borduurt hierop verder, zonder echter het werk van Van den Berg (1949) te vermelden, aan de hand van zijn hypothese over de synthetisering van de bijzin en de werkwoordelijke eindgroep. Die synthetisering, waaronder Duinhoven

het proces verstaat waarbij losse elementen met elkaar verbonden geraken en op de duur als een geheel worden gepresenteerd, uit zich in de twee bovenvermelde tendensen. Hij gaat er dus net als Van den Berg (1949) van uit dat de SVO-volgorde in het Middelnederlands gewoon was en geleidelijk aan is overgegaan naar een SOV-volgorde. Uit de bespreking blijkt dat Duinhoven (1998: 104-107) de SVO-volgorde als de oorspronkelijke Middelnederlandse volgorde beschouwt, waardoor hij aanneemt dat de rode volgorde [AUX+PART] het oudste patroon is (aangezien het deelwoord toen een predicatieve bepaling was, is de kern van de zin het finiete werkwoord). Dat de rode volgorde [AUX+PART] is blijven bestaan, heeft volgens Duinhoven (1998) echter niet zozeer te maken met de conformiteit aan de schrijftaaltraditie, maar wel met de relatieve chronologie van die twee tegengestelde ontwikkelingen: doordat de synthetisering eerst heeft plaatsgevonden, met het deelwoord als kern van de synthetische werkwoordelijke constructie op de laatste positie (resultierend in de rode volgorde [AUX+PART]), en vervolgens pas de voltooiing van de ontwikkeling tot achteropplaatsing van het finiete werkwoord (resultierend in de groene volgorde [PART+AUX]), heeft de rode volgorde [AUX+PART] zich kunnen handhaven. Immers, zo stelt Duinhoven (1998), als de werkwoordelijke constructie als een synthetisch geheel beschouwd wordt, dan is de neiging om het finiete werkwoord achterop te plaatsen minder dwingend. Duinhoven (1998: 98) vermoedt, tot slot, dat de rode volgorde [AUX+PART] sterk aan het verbreiden is. Een aanwijzing daarvoor vindt hij in de hypercorrecte rode volgorde [AUX+PART] die men soms vindt in naamwoordelijk gezegdes.

Het spreekt voor zich dat de analyse van Duinhoven, net als die Van den Berg, op een aantal algemeenheden rust en bovenal gebaseerd is op een toevallig tot stand gekomen set van voorbeeldzinnen. De beide onderzoekers nemen bijvoorbeeld aan dat de SVO-volgorde de oorspronkelijke is, terwijl onafhankelijk corpusonderzoek aangetoond heeft dat de SOV-volgorde in de Middelnederlandse bijzin al heel gewoon, en zelfs de dominante volgorde was (o.a. Gerritsen 1978, Burridge 1993, Coussé 2003, De Schutter 2003; zie ook Van der Horst 1981). Hoe interessant en motiverend de bijdragen van Van den Berg en Duinhoven ook zijn, voorlopig kunnen ze enkel beschouwd worden als hypothese voor toekomstig onderzoek en niet als bewezen feit.

Minder hypothetisch zijn de studies van De Meersman (1990), Burrige (1993) Coussé (2003), Vanacker (1963), Gerritsen (1978), Van Leuvensteijn (1987) en Claes (1992) die zich allemaal op corpusgegevens baseren om uitspraken te doen over vroegere taalstadia van het Nederlands.¹⁷ De Meersman (1990) bestudeert onder andere de volgordevariatie in werkwoordelijke eindgroepen in een reeks stilistisch gediversifieerde, maar toch vooral ambtelijk getinte, Vlaamse en Brabantse teksten uit de 13^{de} tot en met de 16^{de} eeuw. Op basis van dat materiaal komt hij tot de conclusie dat de voorkeur voor de rode volgorde [AUX+PART] tussen de 13^{de} en 16^{de} eeuw afneemt. Dat proces lijkt zich bovendien minder snel voltrokken te hebben in de periferie van het Vlaamse gebied. Problematisch aan de corpusanalyse van De Meersman is dat zijn vaststellingen gebaseerd zijn op erg kleine en onvolledige tellingen (van sommige langere teksten bestudeert hij enkel excerpten) van een veeleer toevallig tot stand gekomen materiaalverzameling. Daardoor is het moeilijk om te evalueren of de vaststellingen van De Meersman (1990) representatief zijn voor die periode en die streek (hetgeen De Meersman zelf ook lijkt te beseffen). Desondanks worden zijn resultaten bevestigd door het onderzoek van Burrige (1993: 117-131) naar Hollands en Brabants proza vanaf het begin van de 14^{de} eeuw tot en met het einde van de 17^{de} eeuw en dat van Coussé (2003) die 1701 bijzinnen uit 13^{de}-, 15^{de}-, 17^{de}-, en 19^{de}-eeuwse Hollandse teksten bestudeerd heeft, eveneens met inbegrip van stilistisch gediversifieerd materiaal¹⁸. Zij komt daarbij tot de vaststelling dat de rode volgorde [AUX+PART] inderdaad afneemt tussen de 13^{de} en 15^{de} eeuw. In tegenstelling tot de verwachting van De Meersman (1990) betekent dat echter niet het geleidelijke, maar definitieve verdwijnen van de rode volgorde [AUX+PART]; het materiaal van Coussé laat namelijk een heropleving van de rode volgorde [AUX+PART] zien vanaf de 17^{de} eeuw. In die zin conflicteren de vaststellingen van Coussé (2003) met die van Burrige (1993), die voor dezelfde periode en dezelfde regio vastgesteld heeft dat de proportie rode volgordes [AUX+PART] 0% benaderde, al moet daar onmiddellijk aan toegevoegd worden dat op basis van het Brabantse materiaal van Burrige (1993) in de loop van de 17^{de} eeuw toch wel een lichte stijging van de rode volgorde [AUX+PART] opgemerkt kan worden.

Tot slot vermelden we hier ook de historisch-diachrone studies van Vanacker (1963), Van Leuvensteijn (1987) en Claes (1992) die zich van de bovenstaande empirische studies onderscheiden doordat ze zich enkel concentreren op één stad,

respectievelijk Aalst, Gouda en Diest. Vanacker (1963: 115-116) merkt in zijn bronnen een absoluut overwicht van de groene volgorde [PART+AUX] (103 vs. 6 rode volgordes [AUX+PART]) in de 15^{de} tot 17^{de} eeuw. Van Leuvensteijn (1987) onderzoekt de woordvolgordevariatie in 110 oorkonden uit de periode 1325-1400 en een dagboekfragment van omstreeks 1575. Hij stelt daarbij vast dat het oorkondenmateriaal een evenwichtige verdeling van de beide varianten in het eerste kwart van de 14^{de} eeuw vertoont en een voorkeur voor de groene volgorde [PART+AUX] in het derde en vierde kwart van de 14^{de} eeuw. Het dagboekfragment dat hij heeft bestudeerd vertoont bijna uitsluitend groene volgordes [PART+AUX]. Claes (1992), die reageert op De Meersmans (1990) vaststelling dat de volgordepreferentie in het *Keurboek van Diest* tussen het begin en het eind van de 15^{de} eeuw omslaat van een voorkeur voor rood [AUX+PART] naar een grote voorkeur voor groen [PART+AUX], merkt in zijn ambtelijk Diestse materiaal, dat vijf bronnen omvat met in totaal 306 eindgroepen die opgebouwd zijn uit een deelwoord en een finiet werkwoord, een stijging van de rode volgordes [AUX+PART], enerzijds van de 14^{de} op de 15^{de} eeuw en, na een lichte daling, van de 15^{de} op de 16^{de} eeuw. Vooral de tweede helft van de 16^{de} eeuw vertoont een spectaculaire stijging van het aantal rode volgordes [AUX+PART]. Die enorme stijging, die drastischer is dan in de rest van Brabant, brengt Claes in verband met de invloed van de Limburgse dialecten, maar hij besluit wijselijk dat daarvoor bijkomend onderzoek nodig is.

Resumerend kunnen we dus zeggen dat de groene [PART+AUX] en rode [AUX+PART] volgorde al eeuwenlang naast mekaar bestaan (maar i.t.t. wat Overduin 1986 beweert, zijn er vooralsnog geen aanwijzingen dat de groene volgorde [PART+AUX] de oudste zou zijn). Hoe de verschillende varianten zich in de loop der eeuwen precies ontwikkeld hebben, is nog grotendeels een open vraag, maar toch lijkt het erop dat na een initiële afname de rode volgorde [AUX+PART] vanaf de 17^{de} eeuw proportioneel toeneemt. Volgens Van der Horst & Van der Horst (1999: 289-290) moet die ontwikkeling geïnterpreteerd worden als een toenemende verbreiding van de standaardtaal, aangezien de meeste dialecten overwegend de groene volgorde [PART+AUX] vertonen (zie paragraaf 3.2.1). Hoe het ook moge zijn, het is duidelijk geworden uit het bovenstaande overzicht, niet in het minst door de contrasterende vaststellingen van een aantal onderzoekers, dat nog heel wat corpusonderzoek, systematisch verzameld en goed gedifferentieerd naar een aantal mogelijke

beïnvloedende intern-linguïstische en extern-linguïstische variabelen, vereist is vooraleer een representatief beeld van de ontwikkeling van de beide varianten geschetst kan worden. Dat daarbij een nauw contact onderhouden zal moeten worden met historici, die op basis van andere dan talige criteria, teksten kunnen situeren in een geografische en sociostilistische dimensie, lijkt vanzelfsprekend. Op die manier zou dan een betrouwbaarder beeld geschapen kunnen worden van de regionale en stilistische diversiteit van de woordschikking in de bijzin en zou een antwoord geboden kunnen worden op de vraag of de opkomst van de eenheidstaal in de 17^{de} eeuw een reductie van de regionale diversiteit teweeg heeft gebracht en er inderdaad voor gezorgd heeft dat de rode volgorde [AUX+PART] kon toenemen. Historische corpusstudies zouden ook de aanvang van de syntactische variatie, in de vroegste nevelen van het Nederlands, met het ontstaan van de perifrastische werkwoordsvormen moeten bestuderen (Quak & Van der Horst 2002: 57-60) en de invloed die Latijnse modellen daarbij gespeeld kunnen hebben (De Meersman 1990: 181, De Schutter 2003: 60-65). Maar ook intern-linguïstische verbanden, zoals de invloed van de positie van de werkwoordelijke groep in de zin op de keuze voor een van de volgordes, zou mogelijk inzicht kunnen verschaffen in de vraag waarom het Nederlands in de loop der eeuwen geëvolueerd is naar een SOV-patroon in de bijzin (in de lijn van Coussé 2003). Voorwaarde daarvoor is wel dat uitgemaakt wordt wat de kern is van de werkwoordelijke eindgroep (Duinhoven 1998 vs. De Meersman 1990, Coussé 2003), wat de verhouding is tussen samenplaatsing en niet-samenplaatsing van de ‘werkwoordelijke’ elementen in de Middelnederlandse bijzin en, fundamentele, vanaf wanneer een clustering van een deelwoord en een hulpwerkwoord als een werkwoordelijke (eind)groep beschouwd mag worden (Van der Wal & Van Bree 1994: 150-153). Ook de invloed van de precieze functie of betekenis van het deelwoord op de woordschikking zou dieper onderzocht moeten worden. Tot slot zou ook nagegaan kunnen worden in hoeverre de ontwikkeling van de beide syntactische varianten gelijkloopt met de ontwikkeling in het Duits, zoals die voor Nürnberg geschetst is door Ebert (1998; zie ook Ebert 1978: 42-43, Takada 1998: 233-234) en het Engels (Pintzuk & Haerberli 2004), en vanaf wanneer en waarom de ontwikkeling is beginnen divergeren (zoals bekend, kent het hedendaagse Duits tegenwoordig enkel nog de groene volgorde [PART+AUX] in de bijzin (SOV), terwijl het Engels geëvolueerd is naar een rode volgorde [AUX+PART] (SVO)). Een antwoord op die vragen is niet alleen van belang voor de historisch taalkundige die geïnteresseerd is in

de historische banden tussen verwante talen, maar ook voor de taalkundige die zich bezighoudt met de volgordevariatie vanuit een puur synchroon perspectief—Ebert (1998: 160-164) vermeldt bijvoorbeeld dat intern-linguïstische factoren zoals de ritmische structuur van belang waren bij de keuze voor een van de woordvolgordes in het zestiende-eeuwse Nürnbergers. Een globale analyse van de woordvolgordevariatie in de hedendaagse Nederlandse werkwoordelijke eindgroep moet daarom zeker rekening houden met de gegevens uit onderzoek naar verwante constructies in verwante talen en moet, indien mogelijk, streven naar een globale verklaring voor die set van verwante talen. Het onderzoek van het taalkundige paradigma dat net van dat comparatieve een van zijn kernpunten heeft gemaakt, komt in de volgende sectie aan bod.

1.3.3 Taalwetenschappelijk onderzoek naar de implementatie van de woordvolgordevariatie in een formeel-autonoom grammaticamodel

1.3.3.1 Generatieve grammatica¹⁹

Nadat de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep door het dialectologische en sociolinguïstische onderzoek (zie 1.3.4) bij een grotere groep linguïsten bekendheid verwierf, zijn ook generatief taalkundigen zich bezig gaan houden met de synchrone beschrijving en verklaring van het fenomeen. Anders dan de dialectologen en sociolinguïsten concentreerden de generatief taalkundigen zich niet op de frequentieverhoudingen van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde in het taalgebruik van verschillende lagen van de bevolking, in verschillende stijlen of in verschillende regio's, maar wel op de syntactische regels in de (universele) mentale grammatica die aan de woordvolgordevarianten ten grondslag liggen.²⁰ Die desinteresse voor taalgebruiksfenomenen vloeit voort uit de doelstelling van de generatieve grammatica om een zicht te krijgen op het aan het taalgebruik ten grondslag liggende taalvermogen, en meer bepaald op de (eindige) set van syntactische regels op basis waarvan alle mogelijke grammaticale zinnen gegenereerd en geïnterpreteerd kunnen worden. Die eindige set regels wordt verondersteld identiek te zijn voor elke moedertaalspreker van het Nederlands en bij uitbreiding voor elke moedertaalspreker van eender welke taal (fylogense), zodat het vanuit een dergelijk oogpunt onzinnig is om verschillen tussen sprekers en groepen en sociale, situationele en ervaringsgerelateerde aspecten van taal te bestuderen. Ondanks het feit dat de syntactische regels van bij de geboorte reeds beschikbaar zijn²¹, hebben linguïsten er

toch geen directe toegang toe, zodat de set van syntactische regels afgeleid moet worden aan de hand van mogelijke grammaticale zinnen (hoe vaak die voorkomen, heeft in dat verband geen enkel belang). Het onderscheid tussen grammaticale en ongrammaticale zinnen kan de generatief taalkundige daarbij intuïtief maken doordat hij, net als iedere moedertaalspreker, over een aangeboren taalvermogen beschikt dat enkel grammaticale zinnen genereert (vgl. Kerstens & Sturm 1979 voor een pleidooi voor het gebruik van intuïties in het generatief-taalkundige onderzoek). Een speciale testcase bij die zoektocht naar de set van syntactische regels zijn allerhande variationele fenomenen. Om een wildgroei van syntactische regels te voorkomen en de syntactische component zo economisch mogelijk te kunnen modelleren, wordt daarbij een onderscheid gemaakt tussen een oppervlaktestructuur en een dieptestructuur, zodat grammaticale vormen die verschillen qua vorm, maar een identieke betekenis hebben (syntactische synonymie) op een zelfde dieptevorm teruggevoerd kunnen worden en identieke grammaticale vormen met een verschillende betekenis (syntactische homonymie) tot een andere onderliggende structuur herleid kunnen worden.²² In het geval van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde komt het erop neer dat het voorkomen van de beide vormen als een oppervlaktefenomeen beschouwd wordt en dat die beide vormen teruggebracht worden naar één gemeenschappelijke onderliggende vorm. Veel aandacht is daarom uitgegaan naar de identificatie van die onderliggende vorm en de transformationele regels die de oppervlaktestructuren genereren. Een niet onbelangrijk aspect bij die zoektocht is het streven naar intertalige overeenkomst, met name de poging om een gemeenschappelijke onderliggende vorm en gemeenschappelijke regels op te stellen voor het Nederlands en de andere (Continentaal) West-Germaanse talen, voornamelijk het Duits.²³

Start- en referentiepunt voor een reeks generatieve analyses van de woordvolgordevariatie in het verbale complex is Evers' (1975) dissertatie over de transformationele cyclus in het Duits en het Nederlands. Zonder zich echt te concentreren op het type intraconstituentenvariatie dat in dit proefschrift centraal staat, suggereert Evers toch dat de groene volgorde [PART+AUX] de onderliggende volgorde is en dat de rode volgorde [AUX+PART] afgeleid wordt van de groene [PART+AUX] door middel van de door hem voorgestelde *V-raising*, een transformatie waarbij het ingebedde deelwoord omhooggetild wordt en geadjungeerd

ter rechter zijde van het matrixwerkwoord (i.c. het hulpwerkwoord).²⁴ De transformatie impliceert dan wel dat de groene volgorde [PART+AUX] niet in de oppervlaktestructuur kan voorkomen. Als dat wel het geval zou zijn, dan zou dat nl. betekenen dat V-raising (ten dele) tegengehouden kan worden. Evers gaat op dat tegenfeitelijke resultaat van de transformatie niet in, maar erkent elders wel dat de groene volgorde [PART+AUX] in de oppervlaktestructuur kan verschijnen (Evers 1975: 51).

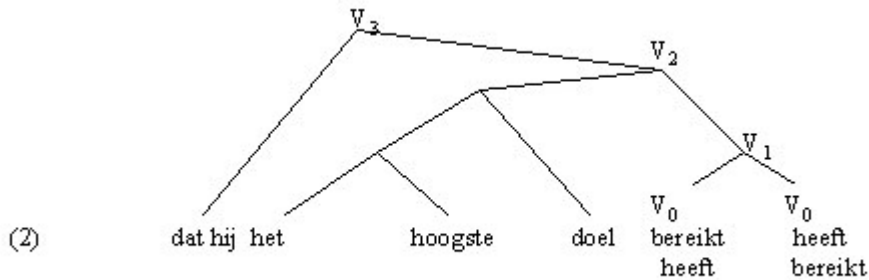
Den Besten & Edmonson (1983) behandelen vanuit een algemene beregeling van de werkwoordsvolgorde in de Continentaal-West-Germaanse talen (Duits, Nederlands en Fries) de volgordekwestie meer in detail. Ze veronderstellen voor die talengroep een algemeen SOV-patroon, omdat het volgens hen het frequentste patroon is²⁵ (het patroon komt nl. niet alleen in bijzinnen voor, maar ook in hoofdzinnen, waar de infiniete werkwoordsvormen achterin de zin staan; de SVO-volgorde in hoofdzinnen wordt in hun analyse verklaard door een verplaatsingsregel die het tijdsdragende verbale element naar de eerste of tweede positie verplaatst). De implicatie van die SOV-assumptie is, in navolging van Greenbergs (1973: 82) universele 16, dat het hulpwerkwoord (V) steeds na het hoofdwerkwoord (O) moet komen (en de groene volgorde [PART+AUX] dus de enige mogelijke volgorde is). Door onder andere een regel van *inversie* te veronderstellen, zorgen Den Besten & Edmonson ervoor dat ook de rode volgorde [AUX+PART], die als een symptoom van een ontluikend taalveranderingsproces beschouwd wordt, mogelijk is.²⁶

Den Besten & Broekhuis (1989) gaan verder op de ingeslagen weg door aan te nemen dat de groene volgorde [PART+AUX] de dieptestructurele volgorde is en de rode [AUX+PART] de geïnverteerde oppervlaktevolverde die middels de transformatie V-raising van de groene volgorde [PART+AUX] afgeleid wordt. Anders dan Evers (1975) beschouwen de auteurs V-raising niet meer als een verplichte transformatie (onder invloed van de vervanging van de transformationele regels door de optionele regel 'move α '), maar als een optioneel proces dat in sommige omstandigheden afgedwongen kan worden. Een probleem voor een dergelijk voorstel, dat erop gericht is om de volgordekwestie in allerlei types eindgroepen te beregelen, blijft uiteraard de optionaliteit in de tweeledige werkwoordelijke eindgroep. De auteurs zien twee oplossingen voor die moeilijkheid: (1) hulphypotheses opstellen die

participia niet tot de syntactische categorie V rekenen, waardoor V-raising niet van toepassing is; (2) V-raising wordt niet afgedwongen, zodat de groene [PART+AUX], niet-geïnverteerde, volgorde de normale volgorde is (maar daardoor wordt het moeilijk om een aantal andere werkwoordelijke eindgroepen die geen groene volgorde [PART+AUX] toelaten, te verklaren). De auteurs geven de voorkeur aan de tweede optie, maar houden zich voor de rest slechts zijdelings met het hier ter discussie staande type variatie bezig.

Sturm (1990) hekelt de voorgaande analyses voor hun contradictorische vaststellingen en hun weinig verklarende karakter. Een verklaring ligt volgens Sturm pas binnen handbereik als er een omvattende theorie zou zijn die uitspraken doet over mogelijke regels van de grammatica's van natuurlijke talen. Daarom bekijkt hij de werkwoordsvariatie vanuit de regeer- en bindtheorie, hetgeen onmiddellijk impliceert dat een transformatie als V-raising als onmogelijk bestempeld moet worden, evenals een eventuele optionalisering van de zijde waar het ingebedde werkwoord geadjungeerd wordt. Hij sluit bovendien een verklaring in termen van een fonologische regel (zie bv. Haegeman & Van Riemsdijk 1986) onmiddellijk uit, omdat die volgens hem gebruikt wordt "als een vergaarbak voor verschijnselen waar men binnen de syntaxis vooralsnog geen raad mee weet" (Sturm 1990: 387). Op basis van het werk van Sassen (1963) wordt de woordvolgordealternatie als syntactisch vrije variatie beschouwd, waardoor de twee volgordes in een verhouding van gestigmatiseerd (groen [PART+AUX]) en prestigieus (rood [AUX+PART]) kunnen staan.²⁷ Vrije variatie houdt op zijn beurt weer in dat er sprake moet zijn van een gemeenschappelijke onderliggende syntactische representatievorm (analoog aan het onderscheid tussen *klank* en *foneem*: klanken die in een verhouding van vrije variatie staan met elkaar, kunnen worden teruggevoerd op dezelfde fonologische representatievorm). De vrije variatie kan daarbij enkel gedefinieerd worden op het oppervlakteniveau, aangezien een definitie op diepteniveau zou betekenen dat syntactische varianten met dezelfde diepterepresentatie eveneens in een verhouding van vrije variatie zouden staan (wat niet het geval is, zie bv. actief vs. passief). Hierdoor komt Sturm tot de conclusie dat de groene [PART+AUX] en rode [AUX+PART] volgorde zich als vrije varianten gedragen, omdat ze in niets te onderscheiden zijn op het oppervlakteniveau. Dat is volgens de standaardopvattingen moeilijk, omdat er een onderscheid gemaakt wordt tussen hulpwerkwoorden (AUX)

en andere types werkwoorden (V); Sturm moet dus eerst, zoals Evers (1975), dat onderscheid opheffen om tot dezelfde oppervlaktestructuur voor de groene [PART+AUX] en rode [AUX+PART] woordvolgorde te kunnen komen: [[AUX][V]]_{VP} en [[V][AUX]]_{VP} worden aldus [[V][V]]_{VP}. Sturm verwerpt bovendien de dieptestructuur van Evers (1975) en neemt aan dat het hulpwerkwoord en hoofdwerkwoord zusterknopen zijn:



Aangezien zowel het hulpwerkwoord als het hoofdwerkwoord V_0 zijn, zoals te zien is in de geconstrueerde voorbeeldzin (2), kan niet uitgemaakt worden welke van de beide het hoofd van de V_1 -constituent is, zodat syntactisch gezien de beide V_0 'en ook omgedraaid kunnen worden en er dus sprake is van syntactische vrije variatie.

De Schutter (1991) draagt verdere argumenten aan voor de analyse van Sturm: doordat de preverbale constituent en de verbale cluster zusterconstituenten op het diepste niveau zijn, kunnen ze fonologisch op elkaar inwerken, hetgeen het effect van de accentuering op de keuze voor een van de woordvolgordes verklaart (zie hoofdstuk 4). Problematisch aan het model-Sturm is echter, zo zegt De Schutter, dat een andere dieptestructuur geconcipieerd moet worden voor de aanwezigheid van partikels en adjectieven in de werkwoordelijke clusters, zoals in een scheidbaar samengesteld werkwoord:

(3) ... [[heeft]_{V0} [[mee]_{V0} [genomen]_{V0}]_{V1}]_{V2}

Daar het partikel *mee* tussen het hulpwerkwoord en het deelwoord staat, en in eerste instantie samengenomen moet worden met *genomen* (waardoor ze samen een V_1 -knoop vormen), kunnen het deelwoord en het hulpwerkwoord niet meer als V_0 -zusterconstituenten fungeren. Dat betekent dat het hiërarchische criterium, dat zegt dat

het meest ingebedde element de kern is (i.c. *heeft*), en het lineariteitscriterium, dat zegt dat wat links staat de bepaling is (i.c. *heeft*), elkaar tegenspreken en dat het model-Sturm aldus zijn aantrekkelijkheid verliest. Voor die gevallen kan echter het model-Evers toegepast worden (rechtsadjunctie en geen semantische scheiding), voor de andere gevallen, met name de gevallen zonder indringing van niet-verbale elementen in de werkwoordelijke eindgroep, acht De Schutter het model-Sturm superieur aan dat van Evers, aangezien dat laatste procedureel ongelukkig is door te veronderstellen dat de V-raising, die de rode volgorde [AUX+PART] in de oppervlaktestructuur veroorzaakt, ten minste indirect op gang zou worden gebracht door de hechte semantische band tussen het deelwoord en de preverbale constituent (in de rode volgorde [AUX+PART] bevindt het hulpwerkwoord zich tussen het deelwoord en de preverbale constituent).

Vanuit de gedachte dat de Germaanse talen op een zelfde manier gestructureerd zijn (althans in de diepstructuur), bestudeert Zwart (1996) in zijn studie over verbale complementatie onder andere de tweeledige werkwoordelijke eindgroep in het Nederlands. Anders dan de andere comparatieve studies neemt Zwart aan dat de Germaanse talen een onderliggende VO-structuur hebben, omdat op die manier de zinsstructurele scheiding tussen Continentaal-West-Germaans en Engels/Noord-Germaans ongedaan gemaakt wordt. De bijzinsvolgorde wordt in die analyse afgeleid door objectverplaatsing. Die aanname lijkt tot gevolg te hebben dat de rode volgorde [AUX+PART] de enige mogelijke volgorde is en dat de groene volgorde [PART+AUX] afgeleid moet worden door linksadjunctie. Zwart behandelt de tweeledige clusters echter op een volstrekt andere manier als hierboven, door de verbale cluster in de lijn van Kayne's possessieve constructies (zie Zwart 1996 voor referenties) te interpreteren. Zo worden possessieve werkwoorden opgevat als samengesteld uit twee hoofden, met name ZIJN en het functionele element VAN (voorbeeld genomen uit Zwart 1996: 247):

- (4) [het boek]_i [_{VP} -- ZIJN [_{VP} -- VAN [_{AgRP} [_{VP} gelezen _{t_i}]]]]
-

Het deelwoord is in die analyse ingebed in de VP onder het VAN-hoofd, dat op zijn beurt geïncorporeerd is in het ZIJN-hoofd (met HEBBEN als resultaat). Doordat het

compositum ZIJN + VAN elk een (V)P projecteert met elk 1 specificierpositie (dus twee in het totaal), zijn er twee specificierposities vrij waarnaar het deelwoord toegelaten wordt te bewegen (er wordt dus sowieso een beweging voorzien voor het deelwoord): een eerste, minder verre, beweging resulteert in de rode volgorde [AUX+PART], de andere in de groene [PART+AUX].

Een laatste generatieve analyse komt van Zuckerman (2001).²⁸ Zuckerman claimt dat echte optionaliteit niet bestaat en dat de selectie tussen alternatieven gebeurt op basis van economieprincipes. In het geval van de groene [PART+AUX] en rode [AUX+PART] volgorde betekent dat volgens Zuckerman dat de groene volgorde [PART+AUX] economischer is dan de rode [AUX+PART], omdat de groene volgorde [PART+AUX] slechts één verplaatsing nodig heeft—het deelwoord wordt verplicht naar links verplaatst in een VP-interne positie—en de rode [AUX+PART] twee—de verplaatsing naar de groene volgorde [PART+AUX] en vervolgens nog een verplaatsing waarbij het hulpwerkwoord links naast het deelwoord komt te staan. Zuckerman zoekt steun voor zijn analyse in het feit dat jonge kinderen meer geneigd zijn om de groene volgorde [PART+AUX] te gebruiken, maar het is duidelijk dat dat gegeven niets zegt over de hoeveelheid en de aard van de verplaatsingen. Opmerkelijk is in ieder geval wel dat Zuckerman (2001) zijn optionaliteits- en economiehypothese tracht te toetsen aan de hand van psycho-experimenteel verkregen resultaten.²⁹

1.3.3.2 *Andere formele grammaticamodellen*

Ook andere formele grammaticamodellen hebben zich, zij het veeleer indirect en zijdelings beziggehouden met de woordvolgordeproblematiek in de werkwoordelijke eindgroep. Een kort overzicht:

- Vanuit *Lexical Functional Grammar* veronderstellen Kaplan & Zaenen (2003) een rechtsvertakkende structuur voor het Nederlands, hetgeen in een rode [AUX+PART] basiswoordvolgorde uitmondt; er wordt echter een uitzonderingsclausule voorzien voor verbale clusters met o.a. een hulpwerkwoord en een deelwoord: daar is een linksvertakkende structuur mogelijk. In plaats van de volgorde aan de top van de cluster te controleren, wordt ervoor geopteerd om functionele voorrangsbependingen toe te voegen aan het lemma van de verschillende types werkwoorden. Die strategie wordt

gehanteerd, omdat zij ervan uitgaan dat de volgorde bepaald wordt door idiosyncratische eigenschappen van lexicale items die van dialect tot dialect, van taal tot taal verschillen. Hierdoor moeten echter opnieuw een aantal extra restricties toegevoegd worden om een aantal mogelijkheden die daardoor ontstaan uit te sluiten.

- Vanuit *Performance Grammar* (Kempen & Harbusch 2003), een psycholinguïstisch gemotiveerd semantisch framework dat bestaat uit verschillende componenten die de hiërarchische en lineaire structuur van zinnen genereren, wordt aan elk lemma in het mentale lexicon een lexicaal frame toegekend met informatie over woordsoort, subcategorisatie, morfologie en in uitzonderlijke gevallen ook positionele kenmerken. Voor het lexicale frame van het hulpwerkwoord betekent dat bijvoorbeeld dat er positioneel kenmerk voorzien wordt dat ervoor zorgt dat het hulpwerkwoord zowel de slot links als de slot rechts van het deelwoord kan innemen.
- *Semantic Syntax* (Seuren 2003), een transformationeel grammaticamodel, ingebed in een algemene architectuur met een link naar cognitie en het medium waarin taal overgedragen wordt, neemt een semantische analyse als input en genereert op basis daarvan een oppervlaktestructuur. In *Semantic Syntax* wordt het Nederlands principieel als rechtsvertakkend beschouwd, waarbij het deelwoord als een kruipfenomeen wordt beschouwd (*creeping*), i.e. het deelwoord kan omhoog kruipen in de structuur (wat de verschillende woordvolgordes mogelijk maakt).
- Vanuit *Head-driven Phrase Structure Grammar* (Bouma & Van Noord 1998) wordt aangenomen dat *complement inheritance* een essentiële operatie is in de analyse van de Nederlandse en Duitse verbale cluster, waarbij o.a. het hulpwerkwoord subcategoriseert voor een potentieel oningevuld verbaal complement en voor alle complementen op de COMPS-lijst van dat verbale complement. Voor de hulpwerkwoorden betekent dat concreet dat ze een deelwoord als complement kunnen nemen, en dat dat zowel links als rechts kan gebeuren.

1.3.3.3 *Computationele toepassingen*

In tegenstelling tot de generatieve-grammaticamodellen die hierboven voorgesteld werden, hebben de computationele NLP-toepassingen niet als doel om de variatie te verklaren of de reële cognitieve processen te identificeren, maar wel om een accurate beschrijving te geven van de verschillende mogelijkheden met als doel een snelle en efficiënte parsing van natuurlijke taal. Een notoir probleem bij een dergelijke automatische analyse van het Nederlands is de parsing van de verbale cluster, niet alleen wegens de variatie, maar ook wegens de zgn. *crossing dependencies*. Het enige ons bekende algoritme dat een NLP-compatibele beschrijving van de verbale cluster tracht te geven, is het *Amazon*-algoritme, dat sinds zijn ontstaan al verschillende gedaanten heeft gekend. De basisdoelstelling, zoals verwoord in Coppen (1987), is een oppervlaktegrammatica te construeren, als module in een groter systeem, die automatisch moet kunnen uitmaken wat verbale clusters zijn en waar ze ophouden. De strategie om dat te realiseren is in de meeste versies geconcentreerd rond het begrip *verwachting*, i.e. sommige hulpwerkwoorden scheppen de verwachting van werkwoorden in een bepaalde vorm die op hun beurt ook weer verwachtingen kunnen scheppen; van zodra er geen verwachtingen meer zijn, is de cluster afgesloten. Voor het deelwoord wordt er rekening mee gehouden dat het op iedere plaats in de cluster kan staan en dus al verwachtingen kan inlossen nog voor ze geschapen zijn (in de groene volgorde [PART+AUX]) (zie ook Van Dreumel & Coppen 2003).

1.3.3.4 *Conclusie*³⁰

Uit het voorgaande overzicht is gebleken dat de generatieve taalkunde, ongeveer 25 jaar na het verschijnen van de eerste studies over volgordevariatie in de Nederlandse werkwoordelijke eindgroep, getracht heeft om de variatie een plaats te geven binnen de generatieve conceptie van het taalsysteem. Daar waar de aandacht in de beginjaren nog vooral ging naar het zoeken van gemeenschappelijke onderliggende structuren en transformaties voor de (Continentaal-) West-Germaanse talen (Evers 1975, Den Besten & Edmonson 1983, Den Besten & Broekhuis 1989), komt de focus vanaf het begin van de jaren '90 meer en meer uitsluitend op de variatie in de tweeledige werkwoordelijke eindgroep met een hulpwerkwoord en een deelwoord te liggen (Sturm 1990, De Schutter 1990, Zuckerman 2001). De graad van succes van die

generatieve analyses wordt over het algemeen—volstrekt theorie-intern—afgemeten aan de hand van drie parameters: “(i) the analysis allows the grammar to be simpler or more restrictive [...]; (ii) the analysis ‘captures generalizations’ that others treat as accidental; and (iii) the analysis makes correct predictions about the well-formedness” (Wasow 2002: 127).³¹ Uit de heterogeniteit van de gepresenteerde analyses blijkt echter dat die zoektocht naar een eenvoudig, algemeen en economisch grammaticaal model ertoe geleid heeft dat er in het generatieve paradigma een zekere rek zit om mogelijke alternatieven te formuleren, zelfs in die mate dat analyses voor een of meerdere aspecten tegenstrijdige uitspraken doen, zodat men zich kan afvragen of de deductie van een abstract model het hoofddoel is, dan wel een beschrijving en verklaring van fenomenen in de natuurlijke taal. Al was de initiële doelstelling van de zoektocht naar een geïnternaliseerde grammatica die abstraheert van performatieve verschillen door verschillende structuren in een hiërarchisch netwerk met elkaar in verband te brengen legitiem, toch lijkt het erop dat de zoektocht naar een zo eenvoudig, omvattend en economisch mogelijk abstract model—wat eenvoudig, omvattend en economisch ook moge betekenen—een doel op zich geworden is en men de band met de natuurlijke taal gaandeweg verloren is. Wat is, bijvoorbeeld, de toetsteen waarmee transformaties of verplaatsingen geëvalueerd worden? Indien die er niet is, wat het geval lijkt te zijn indien de geschetste evolutie van de generatieve analyses op het gebied van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde representatief genoemd mag worden, dan kan men niet anders dan besluiten dat het hoofddoel van de generatieve grammatica een zo eenvoudig, algemeen en economisch model is, los van het feit of dat een goede weergave van de natuurlijke taal is. Zelfs al neemt men nog aan dat het taalvermogen zo eenvoudig mogelijk opgebouwd is—wat nog moet blijken—dan nog hebben de transformaties en verplaatsingen in dat geval geen andere dan theoretische waarde. Een dergelijk abstract model zou dan weliswaar een nuttig object zijn waarmee men vanuit een praktisch computationeel oogpunt de natuurlijke taal op een automatische manier kan verwerken, maar het heeft geen waarde als er geen beschrijving van de humane taalproductie, -perceptie en -verwerving, die in se gebonden zijn aan cognitieve, articulatorische, akoestische en contextuele voorwaarden, mee bedoeld is: “If a theory of language structure is a theory of the representations that are in the minds of the speakers, then it is incumbent on advocates of that theory to suggest how its mental representations manifest themselves in the various ways people use language.

Otherwise, there is little reason to take the psychological claims seriously” (Wasow 2002: 132). In dat verband stelt Wasow (2002: 132) vast dat vroege psycholinguïstische studies er inderdaad van uitgingen dat de taalgebruiker door een onbewuste mentale operatie ging waarbij elementen verplaatst werden, maar dat dergelijke uitgangspunten al gauw als simplistisch beschouwd werden omdat daarvoor te weinig geweten is over het performantiesysteem. Onlosmakelijk verbonden met die discussie is onder andere ook de notie van verklarende kracht binnen een generatief-taalkundig kader en verder de assumptie dat er een strikt onderscheid gemaakt moet worden tussen taalstructuur en taalgebruik, dat het generatief-taalkundige project zich enkel moet concentreren op de taalstructuur (met het daarbij horende primaat van de introspectieve methode) en dat grammaticale theorievorming niet gebaat zou zijn bij de studie van taalgebruiksgegevens.³² Tegen de achtergrond van een aantal inzichten in diverse deelgebieden van de taalkunde, zoals de sociolinguïstiek, psycholinguïstiek en taalverwerving (zie Geeraerts 2003 voor een beknopt overzicht en Tomasello 1998, 2003), zijn een aantal van die assumpties nog moeilijk vol te houden³³ en wordt de noodzaak van een recontextualiserende tendens belicht, i.e. dat de kennis van de taal niet losgekoppeld kan worden van de algemeen-cognitieve, motorische en sociale aspecten van het gebruik die een reële taalgemeenschap van die taal maakt. Dat wil uiteraard niet zeggen dat een aantal uitgangspunten van de generatieve grammatica, zoals de biologische predispositie, genegeerd moet worden, maar wel dat het a priori uitsluiten van zgn. performatieve gegevens in een grammaticaal model artificieel is en een verarming van de taalkundige analyse betekent. In paragraaf 1.4 zullen we de aanpak die wij hier voorstaan, kort schetsen, namelijk een aanpak die een voortdurende, dynamische wisselwerking tussen het taalvermogen en het taalgebruik impliceert en waarbij objectievere methodes, kwantitatieve gegevens en statistische verwerking centraal moeten staan. In de volgende paragraaf bespreken we kort de literatuur die in min of meerdere mate van die principes uitgaat en waarop het huidige onderzoek dan ook geënt zal worden.

1.3.4 Taalgebruiksgeoriënteerd taalwetenschappelijk onderzoek

Het taalgebruiksgeoriënteerd onderzoek naar de volgordevariatie in de werkwoordelijke eindgroep neemt, zoals hierboven al aangegeven werd, aan het eind van de jaren '40, begin van de jaren '50 van de vorige eeuw een aanvang met onder andere de dialectologische studie van Pauwels (1953). Sindsdien is in tal van empirische studies getracht om vat te krijgen op de mechanismen die aan de keuze voor een van de volgordes ten grondslag liggen door de invloed van een resem factoren van prosodische, morfosyntactische, semantische, dialectologische, regionale, stilistische en psycholinguïstische aard te bestuderen (o.a. Meeussen & Vanacker 1951, Pauwels 1953, Sassen 1963, De Schutter 1964, 1976 & 1996, Braecke 1986, Stroop 1970, De Cubber 1986, Haeseryn 1990, Jansen 1990, Pardoën 1991, Swerts 1998, De Sutter et al. 2003 & 2005). Daarbij is in ieder geval duidelijk geworden dat de keuze voor een van de volgordevarianten weliswaar principieel vrij is, maar geenszins een geval van vrije variatie is (zoals Brachin & Vincent (1985) beweren; paragraaf 1.3.1.1), aangezien de keuze bepaald wordt door een (nog niet volledig uitgeklaarde) uiterst opake set van interagerende factoren. Dat men ondanks dat intensieve onderzoek nog niet precies heeft kunnen vaststellen welke set van variabelen de woordvolgorde bepaalt (en vandaar: wat het functionele onderscheid is tussen de twee volgordevarianten), heeft o.i. twee redenen, een methodologische en een inhoudelijke: (i) doordat het vroegere onderzoek gebruik gemaakt heeft van een grote verscheidenheid aan methodes (corpusonderzoek, psycholinguïstisch onderzoek en enquêtering), zich gebaseerd heeft op erg verschillende types taalmateriaal (dialecten, krantenmateriaal, literair materiaal...) en de dataselectie niet altijd even duidelijk gepresenteerd heeft (welk type tweeledige werkwoordelijke eindgroep werd bestudeerd; werden bepaalde subtypes buiten beschouwing gelaten?), bleek het niet onvaak moeilijk om de verschillende onderzoeksresultaten met elkaar te vergelijken en op elkaar te betrekken, eerder geformuleerde hypotheses te verifiëren en vooral om tot een globale linguïstische verklaring van de volgordevariatie te komen. Daarenboven is door het gebrek aan voldoende statistische toetsing vaak onduidelijk hoe betrouwbaar en generaliseerbaar de resultaten zijn en hoe de verschillende factoren zich tot elkaar verhouden in hun effect op de woordvolgorde. Die elementen hebben ervoor gezorgd dat onderzoekers maar matig konden profiteren van de verwezenlijkingen van eerder

onderzoek.³⁴ (ii) De woordvolgordevariatie die hier bestudeerd wordt, is van het intraconstituente type. Zoals eerder gezegd, is hierover nog maar weinig bekend en lijken goed onderzochte verklaringsmechanismen, zoals syntactische complexiteit, pragmatische waarde en ambiguïteitsvermijding die samenhangen met de communicatieve behoeften en/of on line verwerking van taal (Arnold et al. 2000, Wasow 2002), niet toegepast te kunnen worden op dat type van syntactische variatie. De interpretatie van de alternatie tussen de beide volgordes wordt verder nog bemoeilijkt doordat een van de alternerende elementen, het hulpwerkwoord, semantisch min of meer leeg is (dat geldt in ieder geval voor de hulpwerkwoorden die hier bestudeerd zullen worden, *zijn*, *hebben* en *worden*, waar men enkel een aspectuele betekenis kan onderkennen; zie ook De Schutter 1967: 53) en een interpretatie in semantische termen daardoor op de helling komt te staan (zie echter Pardoën 1991 & 1998 die een aspectueel verschil tussen de beide volgordes aanhaalt; zie hoofdstuk 5).

We zouden echter een veel te pessimistisch beeld geven van de stand van zaken in het taalgebruiksgeoriënteerde, empirische onderzoek, indien we zouden beweren dat totnogtoe geen enkele poging ondernomen is om de variatie in de tweeledige eindgroep te verklaren. Uit dat onderzoek komt nl. een viertal dimensies van verklaring naar voren: een sociostilistische en regionale dimensie (bv. Pauwels 1953, Sassen 1963; verder: contextuele dimensie), een prosodische dimensie (bv. De Schutter 1996), een semantische dimensie (bv. Pardoën 1991) en een psycholinguïstische dimensie (bv. Hartsuiker & Westenberg 2000). Tabel 1.1 geeft per dimensie een overzicht van de variabelen die in een of meerdere van de bovenvermelde studies aan bod zijn gekomen en waarvan het effect verklaard is in termen van een van de verklaringdimensies:

Tabel 1.1: Overzicht van de variabelen die gebruikt zijn om vier verklaringprincipes te verifiëren.

Contextuele Dimensie	Prosodische Dimensie	Semantische Dimensie	Psycholinguïstische Dimensie
Regio Register	Structuur deelwoord Inherentie / definieth. preverbale constituent Postverbale groep Lengte middenstuk Locatie accent	Aard van het hulpwerkwoord / deelwoord	Aard vorige cluster

Omdat de bespreking en de verificatie van de vier dimensies in de volgende hoofdstukken ruim aan bod komt, wijden we er hier verder geen aandacht aan. Volledigheidshalve moet hier wel vermeld worden dat naast de variabelen in tabel 1.1 ook een aantal variabelen is onderzocht die niet in een van de dimensies ondergebracht kan worden, omdat het effect en/of de verklaring ervan niet geheel duidelijk is. Zo wordt in Haeseryn (1990: 43-44, 151-153; zie daar ook voor ander onderzoek) onderzocht of de keuze voor een van de woordvolgordes geassocieerd is met het onderscheid tussen de hoofd- en de bijzin. Aangezien we ons in dit proefschrift enkel zullen concentreren op tweeledige werkwoordelijke eindgroepen in de bijzin (zie hoofdstuk 2 voor een volledige omschrijving van de onderzoekscontext) is die variabele voor dit onderzoek niet bruikbaar. Andere variabelen zijn het tempus (Pauwels 1953: 105-115) en de aard van de conjunctie (Pauwels 1953: 118-121). De invloed van de conjunctie hebben we in dit proefschrift niet herhaald, omdat het gebruikte materiaal gecontroleerd is voor het type conjunctie (i.e. er werden enkel bijzinnen bestudeerd die ingeleid worden door het grammaticaal verbindend voegwoord *dat*; zie hoofdstuk 2). De invloed van het tempus op de woordvolgorde komt verderop wel nog ter sprake.

1.3.5 Taalwetenschappelijk onderzoek naar taalnormen

Naast de onderzoeken over het taalgedrag heeft men ook onderzocht wat het oordeel is van taalgebruikers ten opzichte van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde. Niettegenstaande dat deze studie zich enkel zal concentreren op de woordvolgordevariatie in het taalgebruik, en daarom oordelen en attitudes links zal laten liggen, is het vanuit interpretatief opzicht niet onbelangrijk om kort in te gaan op het onderzoek naar de oordelen omtrent de groene [PART+AUX] en de rode [AUX+PART] woordvolgorde, aangezien in het verleden al meermaals is gebleken dat tendensen in het taalgebruik voor een deel begrepen kunnen worden aan de hand van de oordelen die taalgebruikers hebben (zie bv. Knops & Van Hout 1988: 1).

Hoewel het overzicht van de algemene literatuur in paragraaf 1.3.1 geen echt algemene voorkeur voor een van de woordvolgordes aan de oppervlakte bracht, bestaat bij heel wat taalwetenschappers toch de indruk dat de rode volgorde

[AUX+PART] de prestigevariant is (vgl. Sturm 1990 hierboven), getuige het volgende citaat uit de ANS (1997: 1067): “Beide volgordes zijn correct. Het is een wijdverbreid misverstand dat achteraanplaatsing van het deelwoord [de rode volgorde [AUX+PART], GDS] beter zou zijn”. De ANS geeft hiermee onrechtstreeks te kennen dat er in de Nederlandse taalgemeenschap een beleden voorkeur bestaat voor de rode volgorde [AUX+PART]. Het is algemeen bekend dat dergelijke taaloordelen soms met het reële taalgedrag conflicteren, zeker als het om subtiele grammaticale kwesties gaat. Een prachtig voorbeeld daarvan geeft Sassen (1963: 12, ook geciteerd in Haeseryn 1990: 20): “In het tijdschrift ‘Onze Taal’ schrijft iemand: ‘Ik verander in brieven van jonge medewerkers op mijn kantoor ... altijd *meer als* in *meer dan*, ... *bericht worden* in *worden bericht*’. De schrijver vertelt er niet bij waarom hij dat doet, wel schrijft hij zelf even verder: ‘Ik geloof dat ’t verkeerd *geweest is* indertijd de spelling te veranderen’, *geweest is* dus, de veroordeelde volgorde”.

Vanuit die vaststelling en met als doel om de normen van de ANS-redacteuren (1984), en meer bepaald het onderscheid dat de ANS maakt tussen schrijftaal en spreektaal, te vergelijken met die van andere taalgebruikers, is het onderzoek van De Hoop & Smabers (1987) opgevat. De onderzoekers laten daartoe 61 studenten, waarvan er 40 uit de noordelijke provincies van Nederland afkomstig zijn, een oordeel uitspreken over negen grammaticale variabelen, waaronder de variatie in de tweeledige eindgroep. Om na te kunnen gaan of er een verband is tussen woordvolgordepreferentie en register, bieden de onderzoekers de testitems eerst mondeling aan en vervolgens schriftelijk, en vragen de informanten om een persoonlijke en een algemene voorkeur uit te spreken voor een van de varianten. In een derde en laatste opdracht worden de verschillende items nogmaals aangeboden, maar nu in willekeurige volgorde en wordt de informanten gevraagd om de varianten in te schalen op een as met ‘typisch spreektaal’ en ‘typisch schrijftaal’ als polen (vijfpuntenschaal). Het verschil met de eerste twee opdrachten ligt volgens De Hoop & Smabers in het feit dat de eerste peilen naar voorkeuren en de laatste naar een oordeel. De resultaten tonen aan dat de informanten, zowel in de opdracht waar de items mondeling aangeboden werden als in de schriftelijke opdracht, de voorkeur geven aan de groene woordvolgorde [PART+AUX]. Tegen de achtergrond van de regionale afkomst van een groot deel van de informanten is dat resultaat allerm minst verrassend: uit de literatuurbespreking in hoofdstuk 3 zal namelijk blijken dat de

noordelijke provincies een traditioneel ‘groen’ bastion [PART+AUX] zijn.³⁵ Uit het onderzoek blijkt verder nog dat er geen onderscheid is tussen de persoonlijke en de algemene voorkeur. Dat kan o.i. twee dingen betekenen: ofwel is de woordvolgorde voor de informanten een non-issue, waarbij het onderscheid tussen een persoonlijke en een algemene voorkeur sowieso irrelevant is en de voorkeur uitgaat naar de variant waar ze het vaakst mee geconfronteerd worden (i.c. de groene volgorde [PART+AUX]), ofwel is de woordvolgorde wel degelijk een issue en schakelen de informanten hun persoonlijke voorkeur gelijk aan wat ze denken dat de algemene voorkeur is. In het eerste geval is er met andere woorden een projectie van de persoonlijke voorkeur op de algemene voorkeur, in het tweede geval is het net omgekeerd. Als men echter aanneemt dat de schrijftaalvariant over het algemeen de prestigevariant is, i.c. de rode volgorde [AUX+PART] (zie bv. Haeseryn 1990: 41), dan verwacht men dat de informanten, indien ze dat op die manier ervaren, de rode volgorde [AUX+PART] systematisch vaker als schrijftaal zouden inschalen dan de groene volgorde [PART+AUX]. En net dat is niet gebeurd, wat erop zou kunnen wijzen dat de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep een non-issue is en dat de groene woordvolgorde [PART+AUX] voor hen de meer gebruikelijke is. Methodologisch gezien moet uiteraard ook opgemerkt worden dat er wellicht te weinig items aangeboden werden om een representatief beeld te krijgen (er waren slechts twee items met een tweeledige werkwoordelijke eindgroep). Ook over de experimentele omstandigheden is weinig geweten: hoeveel tijd zat er, bijvoorbeeld, tussen de verschillende opdrachten, hoeveel items scheidde de groene [PART+AUX] en rode [AUX+PART] volgorde in de derde opdracht? Dat die vragen erg relevant kunnen zijn bij de interpretatie van onderzoeksresultaten, zal verderop in dit proefschrift nog blijken (hoofdstuk 6).

Haeseryn (1990) heeft eveneens onderzoek gedaan naar oordelen omtrent de groene [PART+AUX] en de rode [AUX+PART] volgorde en heeft daarvoor een lijst van items aan 516 informanten voorgelegd, waaronder 10 items met de rode [AUX+PART] en groene [PART+AUX] volgorde in tweeledige eindgroepen met een deelwoord. Haeseryn vroeg zijn informanten een oordeel te geven op een vijfpuntenschaal met als extremen ‘heel goed’ en ‘heel slecht’. De resultaten uit dat onderzoek laten zien dat de groene volgorde [PART+AUX] globaal genomen (op 2 items na) positiever gewaardeerd wordt (zoals bij De Hoop & Smabers). Wanneer men

die resultaten echter uitsplitst naar regio en beroepsgroep, dan valt het op dat Vlamingen de groene volgorde [PART+AUX] beter vinden dan de rode [AUX+PART] (in vergelijking met de Nederlanders) en dat journalisten de rode volgorde [AUX+PART] beter vinden dan de groene [PART+AUX] (vergeleken met leraren) en, minstens even belangrijk, dat er een overeenkomst lijkt te zijn tussen het oordeel en het gedrag van de informanten (Haeseryn bestudeert in datzelfde onderzoek ook het gedrag van diezelfde informanten in experimentele omstandigheden; zie paragraaf 3.2.1 voor meer uitleg over dat deel van het onderzoek), waardoor Haeseryn (1990: 346) het vermoeden uit dat er wel degelijk sprake is van echt geïnternaliseerde normen.³⁶ Anders dan het onderzoek van De Hoop & Smabers komt in Haeseryns onderzoek wel bij bepaalde groepen sprekers een lichte voorkeur voor de rode volgorde [AUX+PART] naar voren. Die voorkeur vindt men ook bij Swerts & Collier (2000) terug. Zij hebben aan 15 informanten 16 paren van gesproken stimuli aangeboden die variëren volgens type hulpwerkwoord, prosodische structuur en woordvolgorde. De onderzoekers merkten globaal genomen een sterkere voorkeur voor de rode volgorde [AUX+PART] (voor *worden* iets sterker dan voor *zijn*).

Samenvattend kunnen we zeggen dat er op basis van het taalnormengeïnspireerde onderzoek inderdaad wel iets voor te zeggen valt dat de rode volgorde [AUX+PART], op zijn minst voor bepaalde groepen taalgebruikers, prestigieuzer is dan de groene volgorde [PART+AUX], maar van een extreme voorkeur in de volledige taalgemeenschap kan hoegenaamd geen sprake zijn. Waar de rode volgorde [AUX+PART] dat prestige vandaan haalt, is niet geheel duidelijk, al is het niet ondenkbaar dat de algehele dominantie van de groene volgorde [PART+AUX] in de dialecten er wellicht voor iets tussen zit.

1.4 Doelstellingen en uitgangspunten van het onderzoek

Uit het bovenstaande literatuuroverzicht moge alvast duidelijk geworden zijn dat de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep met een deelwoord en een hulpwerkwoord de afgelopen decennia vaak het voorwerp van onderzoek geweest is. De voorliggende studie wil op basis van de verwezenlijkingen van dat onderzoek de zoektocht voortzetten naar het hoe en waarom van die werkwoordelijke variatie in het hedendaagse Belgische en Nederlandse Nederlands,

niet vanuit een taalattitudineel of formeel-taalkundig perspectief, maar vanuit een taalgebruiksgeoriënteerd perspectief. De keuze voor een taalgebruiksgeoriënteerd onderzoek in plaats van een taalattitudineel onderzoek is net als de keuze voor een synchroon perspectief in plaats van een diachroon perspectief door de complementariteit van de benaderingen arbitrair³⁷, de keuze voor een taalgebruiksgeoriënteerd onderzoek versus een formeel-taalkundig onderzoek is dat in veel mindere mate. In paragraaf 1.3.3.4 hebben we trachten toe te lichten dat we die laatste benadering minder geschikt achten, onder meer omdat we het strikte onderscheid tussen taalstructuur en taalgebruik en het primaat van de taalstructuurstudie (en daaruit volgend: het primaat van de introspectieve methode) tegen de achtergrond van recente ontwikkelingen in de taalkunde als onhoudbaar beschouwen.

Voor onze zoektocht naar het hoe en waarom van de werkwoordelijke variatie in tweeledige werkwoordelijke eindgroepen met een hulpwerkwoord en een deelwoord gaan we ervan uit dat de variatie geen inbreuk maakt op het Humboldtiaanse isomorfieprincipe, i.e. “[...] the old principle that the natural condition of a language is to preserve one form for one meaning, and one meaning for one form” (Bolinger 1977: x-xi) of, in een ruimere interpretatie, *dat een taal geen alternatieven ontwikkelt die al eeuwen naast mekaar bestaan zonder dat er een functioneel onderscheid mee gemoeid is*. Vooraleer de vraag naar het waarom van de coëxistentie van de syntactische alternatieven behandeld kan worden, zullen we ons eerst afvragen wat ervoor zorgt dat taalgebruikers nu eens de groene [PART+AUX] en dan weer de rode [AUX+PART] volgorde kiezen. Die vraag proberen we te beantwoorden door aan de hand van reëel geproduceerde eindgroepen uit twee zorgvuldig gecompileerde corpora de factoren te identificeren die de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde in het hedendaagse taalgebruik van het Nederlands bepalen. Met die basisdoelstelling, dat basisuitgangspunt en de voorgestelde methode verschilt de voorliggende studie nauwelijks van het gros van de taalgedragsonderzoeken die in paragraaf 1.3.4 gepresenteerd zijn, maar dat betekent hoegenaamd niet dat we met deze studie geen bijdrage zouden willen leveren aan het onderzoek naar de volgordevariatie in de tweeledige werkwoordelijke eindgroep in het bijzonder en aan het

taalwetenschappelijke onderzoek in het algemeen. Die bijdrage situeren we op vier vlakken:

- Allereerst is deze studie er uiteraard *op inhoudelijk vlak* op gericht om de kennis over de woordvolgordevariatie te vergroten. Om dat te verwezenlijken zal deze studie in eerste instantie geënt worden op de verwezenlijkingen van vroeger onderzoek: de resultaten van vroeger onderzoek zullen aan de hand van ons materiaal geverifieerd worden, de verschilpunten en gelijkenissen die daaruit voortvloeien, zullen aangestipt en geïnterpreteerd worden. In een tweede, innovatieve stap zal kritisch nagedacht worden over de manier waarop in het verleden linguïstische hypothesen geoperationaliseerd zijn, zullen voorstellen tot verbetering aangereikt worden, maar er zullen ook nieuwe hypothesen voorgesteld worden en nieuwe en meer diepgravende analyses uitgevoerd worden met het oog op de vervollediging, verfijning en verdieping van de bestaande kennis over de specifieke werking van de syntactische variatie in het bijzonder en de werking van de taal in het algemeen. Die verificatie- en innovatieprocedure zal, tot slot, resulteren in een nieuw overkoepelend model dat de coëxistentie van de groene [PART+AUX] en de rode [AUX+PART] volgorde wil verklaren.
- In de tweede plaats willen we *op methodologisch vlak* met deze studie aantonen dat een empirische studie gebaat is met een doorgedreven corpusanalyse indien men een goed inzicht wil krijgen in het voorkomen en de frequentie van een bepaald taalgedragsfenomeen en de factoren die er een invloed op uitoefenen. Een doorgedreven corpusanalyse betekent niet alleen dat men reëel taalgedrag bestudeert aan de hand van corpusgegevens, het betekent eveneens dat men de relevante taalgegevens op een zorgvuldige manier selecteert, op een objectieve en systematische manier volledig analyseert, met speciale aandacht voor grensgevallen, operationalisering van linguïstische hypothesen en andere designerrelateerde kwesties, en dat men de resultaten van de analyses kadert tegen de achtergrond van de structuur en inhoud van de empirische fundamenten. De twee cruciale voordelen van een dergelijke doorgedreven aandacht voor de methodologische fundamenten van wetenschappelijk onderzoek zijn een verhoogde interpreteerbaarheid van de

resultaten en de mogelijkheid tot repliceerbaarheid in toekomstig onderzoek (zie de opmerkingen die we in 1.3.4 maakten over het vroegere onderzoek).

- In de derde plaats wil de huidige studie *op analytisch vlak* een pleidooi houden voor het systematische gebruik van een set van inferentieel-statistische technieken om de betrouwbaarheid en representativiteit van de onderzoeksresultaten na te gaan. We willen daarbij verder gaan dan de traditionele chi-kwadraatanalyses, die nagaan *of* een factor een significante invloed heeft op het woordvolgordegedrag, door andere bivariate, trivariate en multivariate statistische tests te gebruiken die een antwoord kunnen bieden op de vraag *hoe sterk* het effect van een bepaalde factor is op de woordvolgorde; wat de *verbanden* zijn tussen de verschillende factoren onderling (interactie en redundantie); wat het *gecombineerde effect* is van alle factoren op de woordvolgorde; in welke mate de woordvolgorde *voorspeld* kan worden op basis van de geselecteerde factoren. Daarbij zal veel aandacht besteed worden aan de verschillende voorwaarden waaraan voldaan moet worden om een betrouwbaar resultaat en een geldige interpretatie te verkrijgen.³⁸
- Tot slot wensen we met ons onderzoek ook *op theoretisch vlak* iets bij te dragen en wel door aan te tonen dat het onderzoek naar variationele fenomenen gebaat is met de studie van zowel taalinterne als taalexterne factoren en dat beide types factoren tegelijk evenwaardig kunnen bijdragen tot de verklaring van de verspreiding van de varianten. Daarnaast wil deze studie ook een voorbeeld geven hoe een taalgebruiksgebaseerde en emergente benadering van de taal en de grammatica concreet gestalte kan krijgen (Barlow & Kemmer 2000, Bybee & Hopper 2001). Dat wil zeggen dat we, in tegenstelling tot de uitgangspunten van de formele grammaticamodellen (1.3.3), ervan uitgaan dat er geen strikt onderscheid gemaakt kan worden tussen taalstructuur en taalgebruik en dat er een voortdurende en dynamische wisselwerking is tussen taalstructuur en taalgebruik: mentale representaties van taalstructuur worden enerzijds beschouwd als een voorlopige stand van zaken die gevoelig zijn voor en constant adapteren aan het taalgebruik, anderzijds bepaalt de manier waarop gegevens opgeslagen zijn in het geheugen (taalstructuur) de manier waarop we er gebruik van maken (taalgebruik). We nemen met andere woorden aan dat er een voortdurend proces van structurering en herstructurering aan de gang is,

waarbij concrete uitingen afstammen van het abstracte systeem, maar tegelijkertijd het linguïstische systeem herdefiniëren.

1.5 Opbouw van de studie

Nu de doelstellingen van de studie en de uitgangspunten waarvan we vertrekken duidelijk zijn, sluiten we dit hoofdstuk af met een kort overzicht van de manier waarop het materiaal in de volgende hoofdstukken georganiseerd is. Tegen de achtergrond van het feit dat een zorgvuldig bronnenbeheer een *conditio sine qua non* is voor een accurate interpretatie van de verkregen resultaten, zal hoofdstuk 2, dat het inleidende en voorbereidende gedeelte van dit proefschrift afsluit, gewijd zijn aan de methodologische grondslagen van het onderzoek, met name de gebruikte corpora, de selectievoorwaarden waaraan de corpusgegevens moeten voldoen en de statistische verwerking van de gegevens. Vanaf hoofdstuk 3 tot en met hoofdstuk 6 komt het eigenlijke empirische onderzoek aan bod. We wijden telkens een hoofdstuk aan een van de dimensies die we in paragraaf 1.3.4 onderscheiden hebben: zo zal in hoofdstuk 3 de invloed van de regionale afkomst en de stilistische omstandigheden op de keuze voor de woordvolgorde bestudeerd worden (de contextuele dimensie), in hoofdstuk 4 komt het effect van de accentdistributie in en rond de werkwoordelijke eindgroep aan bod (de prosodische dimensie), in hoofdstuk 5 wordt ingezoomd op de invloed van de semantiek (de semantische dimensie) en in hoofdstuk 6 wordt tot slot (onder meer) de rol onderzocht die een voorafgaande werkwoordelijke eindgroep speelt bij de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde (de psycholinguïstische dimensie). In tabel 1.2 wordt een overzicht gegeven van de verschillende dimensies in het onderzoek en de basisfactoren³⁹ die er besproken zullen worden:

Tabel 1.2: Overzicht van de dimensies en basisfactoren die in dit proefschrift aan bod zullen komen.

Hoofdstuk 3	Hoofdstuk 4	Hoofdstuk 5	Hoofdstuk 6
Contextuele Dimensie	Prosodische Dimensie	Semantische Dimensie	Psycholinguïstische Dimensie
Regio Register	Structuur deelwoord Inherentie/definietheid preverbale constituent Postverbale groep Lengte middenstuk Locatie accent	Aard van het deelwoord Aard van het hulpwerkwoord	Aard vorige cluster Frequentie van het deelwoord

Elk hoofdstuk van het empirische gedeelte vangt (voor zover mogelijk) aan met een status quaestionis en test vervolgens aan de hand van de corpusgegevens in de dataset een aantal hypothesen die uit het vroegere onderzoek naar voren gekomen zijn en een aantal nieuwe hypothesen. De analyses die in die hoofdstukken uitgevoerd zullen worden, zijn voornamelijk bedoeld om een antwoord te bieden op de vraag of een bepaalde factor een invloed heeft op de keuze voor een van de woordvolgordes en, zo ja, hoe sterk die invloed dan wel is. Een globale statistische analyse, waarbij het effect van de verschillende statistisch significante factoren uit de hoofdstukken 3 tot en met 6 tegenover elkaar afgewogen wordt en waarbij nagegaan wordt in welke mate die set van factoren in staat is om de totaal geobserveerde variatie te verklaren en te voorspellen, komt in hoofdstuk 7 aan bod. Aansluitend bij en zich baserend op de globale statistische analyse wordt in hoofdstuk 7 ook naar een globale linguïstische verklaring gezocht voor de coëxistentie van de groene [PART+AUX] en de rode [AUX+PART] woordvolgorde. Hoofdstuk 8, tot slot, sluit deze studie af door de verwezenlijkingen van de studie samen te vatten en mogelijke perspectieven voor toekomstig onderzoek aan te wijzen.

Noten:

1. Naast de term *werkwoordelijke eindgroep* zal in deze studie ook gebruik gemaakt worden van de volgende (synonieme) termen: *eindgroep*, *verbale* of *werkwoordelijke cluster* en *cluster*.

2. Opgemerkt moet worden dat niet altijd even duidelijk is welke constituenten er in de polen terecht kunnen komen. In een hoofdzin is de eerste pool volgens de ANS (1997) gereserveerd voor het finiete werkwoord, terwijl voor Jansen (1979: 85) het subject en het finiete werkwoord samen de eerste pool bezetten: “the subject and finite verb occupy positions in front of the sentence in main clauses. They are the left-hand braces in the sentence.” Aangezien we ons in deze studie op de tweede pool zullen focussen, gaan we op dit twistpunt niet dieper in. Relevanter is de indeling die De Schutter (1976) hanteert. Hij stopt in tegenstelling tot de ANS ook inherente, niet-verbale zinsdelen, zoals het naamwoordelijk deel van het naamwoordelijk gezegde, in de tweede pool (hij noemt dat het vierde zinsstuk) terwijl die voor de ANS tot het middenstuk behoren. In hoofdstuk 4 zullen we nog uitgebreid op die indelingskwestie terugkomen, voorlopig kunnen we volstaan met de algemeen aanvaarde indeling van de ANS (1997). Een verhelderende inleiding in aard en ontstaan van de tangconstructie en het gebruik ervan in verschillende theoretische paradigmata vindt men overigens in Van Canegem-Ardijns (2004: 19-55). In diezelfde studie (Van Canegem-Ardijns 2004: 57) wordt ook een alternatieve invulling van het concept tweede pool gepresenteerd (met name de functionalistische idee dat de tweede pool de meest inherente tweede relatum van het relator-hoofd (de eerste pool) is. In de voorbeeldzinnen zou dat betekenen dat *worden* de tweede pool vormt, *dat* de eerste pool en het deelwoord respectievelijk tot het middenstuk (groene volgorde) of de laatste zinsplaats (rode volgorde) behoort (zie bv. Van Langendonck 2003).
3. De referentiegrammatica ANS (1997: 109-111) maakt een onderscheid tussen een voltooid deelwoord en een passief deelwoord. Aangezien er geen formeel verschil bestaat tussen de beide types en het verschil voorlopig niet relevant is voor het onderzoek—het verschil bestaat erin dat het eerste type deelwoord in actieve zinnen gebruikt wordt en het tweede type in passieve zinnen—zullen we *deelwoord* gebruiken als overkoepelende term voor het voltooid deelwoord en het passief deelwoord (maar niet voor het tegenwoordig deelwoord). Wanneer een specifiek type deelwoord benoemd moet worden, zullen we terugrijpen naar de termen *voltooid* en *passief deelwoord*.
4. Alle voorbeelden in deze studie zijn, tenzij anders vermeld, geattesteerde zinnen of paradigmatische varianten van geattesteerde zinnen. Paradigmatische varianten worden enkel gecreëerd in de context van geattesteerde voorbeelden. Na elke geattesteerd voorbeeld wordt de corpuscomponent waaraan het onttrokken is (in dit voorbeeld: *Het Laatste Nieuws*) tussen haakjes vermeld. Zie hoofdstuk 2 voor de andere corpuscomponenten.
5. We zullen in het vervolg zowel gebruik maken van de traditionele terminologie *groene* en *rode woordvolgorde* als van de formele notatie [PART+AUX] en [AUX+PART] om de beide syntactische alternatieven te benoemen.
6. Het moge duidelijk zijn dat het onderscheid tussen een grammatica en taaladviesliteratuur niet altijd even duidelijk is. Over het algemeen hebben we het volgende criterium gebruikt om een onderscheid te maken: indien het boek een min of meer systematische beschrijving geeft van (ten minste) de syntaxis van het Nederlands, dan werd het als *grammatica* beschouwd; in het andere

geval werd het tot de categorie van de *taaladviesliteratuur* gerekend. Het criterium bleek in de praktijk werkbaar genoeg om eenduidige beslissingen te kunnen nemen.

7. Booij & van Santen (1998: 164) vinden de term *scheidbaar samengesteld* slecht gekozen, wegens intern tegenstrijdig, en verkiezen daarom *samenkoppeling*. Omdat de keuze voor een van de termen geen enkele consequentie heeft voor de analyses, noch terminologische verwarring kan veroorzaken, houden we het in deze studie bij de meest gebruikte term *scheidbaar samengesteld werkwoord*.
8. Ons overzicht is gebaseerd op Klein (1999).
9. We nemen aan dat Belgische uitgevers zich voornamelijk richten op een Belgisch doelpubliek en Nederlandse uitgevers op een Nederlands doelpubliek. Het spreekt echter voor zich dat dat onderscheid in de praktijk, en wellicht nu meer dan vroeger, voor Vlaanderen meer dan voor Nederland, genuanceerd moet worden.
10. Zo hebben we bijvoorbeeld de verschillende jaargangen van *Onze taal*, *Van Dale taalbrief*, *Tekst[blad]*, *Nu nog!Nederlands van nu*, *Over taal*, *Taalbeheersing in de praktijk* en *Taalbeheersing in de administratie* buiten beschouwing gelaten.
11. Publicaties waarvan niet uitgemaakt kon worden wanneer ze verschenen zijn (gemarkeerd in de literatuurlijst door s.d.), werden uit de analyse geweerd (n=3).
12. We laten hiervoor *Taalbaak* buiten beschouwing.
13. Van der Horst (2000: 94, 2002: 549-550) vermoedt dat Charivarius rond 1940 of kort voor 1940 de norm introduceerde dat de rode volgorde [AUX+PART] beter is dan de groene [PART+AUX].
14. In tegenstelling tot wat Duinhoven (1998: 97) beweert, beveelt enkel het stijlboek van De Volkskrant de rode volgorde aan. Het Stijlboek van de VRT beveelt de groene volgorde [PART+AUX] aan, de stijlboeken van Trouw, NRC Handelsblad en De Standaard laten de keuze vrij of vermelden de variatie niet.
15. Zie ook Michels (1957: 107) over Vondels taaloordeel omtrent de woordvolgordevariatie en Sassen (1963: 14-15) die een aantal (veelal impressionistische) tellingen in oudere werken heeft uitgevoerd.
16. Van den Berg (1949) is niet de enige die die twee tendensen heeft opgemerkt: Gerritsen (1978) beschrijft, gedeeltelijk aan de hand van frequentietellingen, de ontwikkeling van de volgordepatronen vanaf het Indo-europees en Oer-Germaans over het Oud- en Middelnederlands tot het Nieuw-Nederlands. Zij gaat daarbij ook in op de woordvolgorde in verwante Germaanse talen. Zie ook Van der Horst (1981: 38-48), Burridge (1993: 43 e.v.) en voor een ander perspectief op de volgordeverschuivingen De Schutter (2003). Over de ontwikkeling van het deelwoord en het

ontstaan van perifraistische werkwoordstijden, zie o.a. Stoett (1909: 204-205), Duinhoven (1985, 1997 & 1998), Quak & Van der Horst (2002: 57-58).

17. Hiermee zijn zeker niet alle onderzoeken over de werkwoordsvolgorde in de oudere taalfasen van het Nederlands vermeld (bv. Gerritsen 1978). Voor een aantal andere studies, zie bv. De Meersman (1990).
18. Jammer genoeg splitst Coussé (2003) haar resultaten verderop niet uit volgens register.
19. De term ‘generatief’ wordt hier overkoepelend gebruikt om de set van intergerelateerde theorieën aan te duiden die een aantal basisprincipes en -assumpties met elkaar gemeen hebben (o.a. Principles & Parameters, Regeer- & Bindtheorie). Dat betekent niet dat alle veronderstellingen die in de bespreking hierna volgen voor al die (deel)theorieën gelden.
20. Merk op dat de generatief taalkundigen als doel hebben een (abstract) model te ontwikkelen dat het taalvermogen van de taalgebruiker representeert (een economische, overkoepelende beschrijving met voorspellende waarde) en hoegenaamd geen uitspraken wensen te doen over performancegerelateerde psychologische processen.
21. Een reden voor deze assumptie ziet de generatieve taalkunde in het feit dat kinderen op basis van deficiënte en onvolledige gegevens uit hun omgeving bijzonder snel hun moedertaal verwerven. De aangeboren capaciteit tot taalverwerving neemt de vorm aan van een taalverwervingsmechanisme (het zgn. *Language Acquisition Device* of *LAD*), een systeem van principes dat het kind leidt bij het herkennen van talige eigenschappen en het construeren van het regelsysteem.
22. Het onderscheid tussen oppervlakte- en dieptestructuur is in recente versies van het generatieve model afgeschaft (Chomsky 1995).
23. Algemeen wordt aangenomen dat in het Duits enkel de groene volgorde [PART+AUX] mogelijk is. Dat geldt weliswaar voor het Hochdeutsch, maar niet voor sommige dialecten (zie hoofdstuk 3). Dat gegeven wordt door geen enkel van de hieronder besproken analyses in rekening gebracht.
24. Een aantal aspecten van V-raising, zoals bv. het snoeien van de dominerende S, wordt hier niet besproken.
25. De auteurs geven daarbij niet aan waar ze dat vandaan halen, zodat vermoed kan worden dat met frequentie ‘intuïtief aangevoelde’ frequentie bedoeld is.
26. Sturm (1990: 387) merkt op dat Den Besten & Edmonson even verderop de coëxistentie van V-raising en inversie afschaffen, omdat die beide regels altijd samen voorkomen. Daarmee suggereren Den Besten & Edmonson dat enkel de groene volgorde [PART+AUX] in de oppervlaktestructuur voorkomt.

27. Sturm negeert wel een aantal andere factoren waarvan vroeger onderzoek heeft aangetoond dat ze relevant zijn (zoals terecht opgemerkt wordt in De Schutter 1990). Sommige onderzoeksresultaten, zoals die van Pauwels (1953), worden daarenboven nogal snel afgedaan als onbetrouwbaar. De methodologische bezwaren die Sturm maakt tegen het onderzoek van Pauwels zijn weliswaar terecht, maar dat betekent niet dat het werk volledig genegeerd moet worden. De enige manier om uit te maken wat de invloed geweest is van bepaalde methodologische keuzes van Pauwels is door empirisch na te gaan in hoeverre de resultaten behouden blijven na verificatie met inbegrip van eliminatie van de betreffende methodologische bezwaren.
28. Zie verder ook nog Evers (2003), waarin de variatie in de tweeledige werkwoordelijke eindgroep beschreven wordt aan de hand van de eigenschappen van een *zgn. cluster creeper*.
29. Daarbij kan men zich afvragen hoe een dergelijke empirische aanpak gerijmd kan worden met de voor het overige klassiek-generatieve aanpak die Zuckerman hanteert. Daarnaast moet ook opgemerkt worden dat niet altijd even zorgvuldig omgegaan wordt met de methodologische en analytische eisen die gepaard gaan met empirisch onderzoek. Zo wordt bijvoorbeeld niets gezegd over de mate van controle bij het creëren van de stimuli (frequentie, lengte, thematische argumenten, type werkwoorden...) en wordt er wel erg creatief omgesprongen met de resultaten van vorig onderzoek: zo baseert Zuckerman zich op het werk van Pardoën (1991), maar weigert hij de resultaten van Pardoën's onderzoek aan een kritische analyse te onderwerpen en, wat erger is, negeert hij heel wat van het andere onderzoek dat de laatste decennia uitgevoerd is (of hij noemt het achterhaald, zonder een ernstig argument aan te bieden waarom dat dan wel zo zou zijn), wat er bijvoorbeeld toe leidt dat hij aanneemt dat er geen regionale of stilistische variatie in het spel is (terwijl gedegen empirisch onderzoek nu net heeft aangetoond dat dat wel het geval is). De resultaten van zijn eigen analyses, die inherent gradueel van aard zijn, interpreteert hij categorisch: zo zegt hij van sprekers die een grotere voorkeur hebben voor de ene dan voor de andere volgorde dat dat bewijst dat binnen een I-taal (het taalsysteem) geen optionaliteit bestaat (zie ook noot 37).
30. We focussen ons voor deze conclusie uitsluitend op de generatieve grammatica. De opmerkingen en vragen die daarbij gesteld zullen worden, gelden echter *grosso modo* ook voor de andere formele paradigma's en de computationele toepassingen, voor zover zij met hun analyse een representatie van een homogene interne grammatica nastreven.
31. Echter, "It is hard to assess simplicity or restrictiveness without precise characterizations of these notions. [...] in a framework lacking the sort of explicit description required for useful evaluation of its simplicity and restrictiveness" (Wasow 2002: 127).
32. Een interessante discussie die enkele van de hier genoemde aspecten behandelt, ontspon zich aan het begin van de jaren '80 in het tijdschrift *De nieuwe taalgids*: zie achtereenvolgens Kerstens & Sturm (1979), Van der Lubbe (1980), Verhoeven (1981a), Sturm & Kerstens (1981) en Verhoeven (1981b).

33. Hierbij moet opgemerkt worden dat het primaat van de intuïtie in het generatieve paradigma niet absoluut is. Zo gebruikt Zuckerman (2001) in zijn doctorale proefschrift de resultaten van een aantal experimenten als argument voor een bepaalde (typisch generatief-taalkundige) analyse. De manier waarop met die resultaten echter omgegaan wordt, zullen we in 1.4 afzetten tegen een doorgedreven empirische manier van analyseren die wij hier voorstaan.
34. Met die methodologische en analytische bezwaren moet in ieder geval rekening gehouden worden bij de bespreking van de resultaten van vroeger onderzoek (vanaf hoofdstuk 3). Gelijkenissen en verschilpunten tussen het voorliggende proefschrift en eerder onderzoek zullen dan ook slechts conditioneel geduid kunnen worden.
35. Bij nadere beschouwing van de tabellen in De Hoop & Smabers (1987: 298) blijkt echter dat er voor de gesproken taal een significant verschil is (eigen analyse) tussen de verdeling van de groene [PART+AUX] en rode [AUX+PART] volgorde in item 19 en 20. In het eerste item wordt de groene volgorde [PART+AUX] geprefereerd, in het tweede de rode [AUX+PART]. De voorkeur voor de groene volgorde [PART+AUX] blijkt dus niet zo absoluut te zijn als De Hoop & Smabers doen uitschijnen. Indien men de resultaten voor de gesproken en geschreven taal optelt (er is geen significant verschil tussen die twee modi), dan valt het significante verschil in distributie tussen 19 en 20 weg.
36. In hoofdstuk 3 wordt die interpretatie aan een kritisch licht onderworpen.
37. Dat betekent echter niet dat het overzicht van het historisch-diachrone onderzoek (1.3.2) en het onderzoek naar de taalnormen (1.3.5) voor niets geweest is. Vooral in hoofdstuk 3, waar de contextuele effecten op de woordvolgorde ter sprake komen, zal nog veelvuldig verwezen worden naar het literatuuroverzicht in die beide paragrafen.
38. Dat in het verleden al te slordig omgesprongen is met de selectie en het gebruik van statistische technieken kan treffend geïllustreerd worden aan de hand van het onderzoek van Zuckerman (2001), waarnaar hierboven verwezen is. Zuckerman (2001) gebruikt voor de statistische analyse van de gegevens zowel de t-test (tabel 4, 5 en 7, p. 82-84), de correlatiecoëfficiënt en de chi-kwadraattoets (tabel 8, p. 87). In sommige gevallen gebruikt hij zelfs totaal geen statistische toetsen (tabel 1, 2, 3 en 6, p. 80-81, 83), zonder daarvoor een specifieke reden op te geven. Het gebruik van de chi-kwadraattest in het onderzoek van Zuckerman is onproblematisch, maar bij het gebruik van de t-test en de correlatiecoëfficiënt kan toch heel wat opgemerkt worden: de t-test veronderstelt namelijk een normaal (klokvormig) verdeelde dataset en om een correlatie-analyse uit te voeren heb je minstens twee variabelen op de intervalschaal nodig. Op basis van de informatie die de lezer geboden wordt, kan men twijfelen of aan die voorwaarden voldaan is, waardoor de resultaten van Zuckermans onderzoek onbetrouwbaar lijken.
39. Sommige van die basisfactoren zullen op verschillende manieren geoperationaliseerd worden (zie de doelstellingen in paragraaf 1.4).

Hoofdstuk 2

Methodologische grondslagen van het onderzoek

2.1 Inleiding

In dit hoofdstuk worden de empirische fundamenten voorgesteld van ons onderzoek naar de volgordevariatie in tweeledige werkwoordelijke eindgroepen met een deelwoord en een hulpwerkwoord. Die fundamenten bestaan uit reële, spontaan geproduceerde taaluitingen die uit twee verschillende corpora onttrokken zijn. Het gebruik van geattesteerde gegevens in een onderzoek naar variatie in het hedendaagse Nederlandse taalgebruik is oncontroversieel en, wat meer is, dergelijke gegevens worden zelfs als de enige geschikte beschouwd. In het verleden is nl. al meermaals gebleken dat het bijzonder moeilijk is om zich op intuïtieve basis een goed beeld te vormen van de verdeling en verspreiding van syntactische varianten, laat staan van de onderliggende motieven: de processen die er werkzaam zijn, zijn zo subtiel dat ze met ‘het blote oog’ niet zichtbaar zijn en zelfs als ze al voor ‘een geoefend oog’ enigszins zichtbaar zouden zijn, dan nog kan er geen zekerheid bestaan over de realiteitswaarde ervan. Naast introspectie biedt ook enquêtering in die omstandigheden geen uitkomst, omdat het gevaar bestaat dat informanten zich laten leiden door hun normgevoelens (zie bv. Swerts & Collier 2000) of door wat ze verwachten dat de interviewer van hen wil horen (de zgn. *observer’s paradox*) en/of vaak geen precies idee hebben van wat ze precies allemaal gebruiken: “what people think they do with words is not necessarily the same as what they actually do, in the sense that our conscious

awareness of the flexibility with which we use the resources of the language may well be rather restricted” (Geeraerts, Grondelaers & Bakema 1994: 17).

De aandacht in dit hoofdstuk gaat uit naar de corpora waaraan we de relevante taaluitingen willen extraheren (paragraaf 2.2), de selectierestricties waaraan de data moeten voldoen om als relevant beschouwd te worden (paragraaf 2.3), de zoekopdracht waarmee we de gegevens in de corpora opgezocht hebben (paragraaf 2.4), de controle die we uitgevoerd hebben om na te gaan of alle gegevens in de dataset relevant zijn (paragraaf 2.5), de codering en de statistische verwerking van de corpusgegevens (paragraaf 2.6).

2.2 Selectie van het materiaal: inhoud en structuur van de gebruikte corpora

Voor het onderhavige onderzoek maken we enerzijds gebruik van het digitale, niet-geannoteerde *CONDIV-corpus geschreven Nederlands* (Grondelaers et al. 2000; verder: CONDIV), dat in totaal ca. 45 miljoen woordtokens bevat, en anderzijds van de vijfde release van het *Corpus Gesproken Nederlands* (enkel het gedeelte dat met woordsoortinformatie verrijkt is, i.e. ca. 3,7 miljoen woordtokens; Oostdijk 2000; verder: CGN (R5)). Beide corpora bevatten spontaan tot stand gekomen, reëel taalgebruik en zijn geconstrueerd volgens een geografische en een stilistische dimensie¹.

Tabel 2.1: Structuur en inhoud van het CONDIV-corpus geschreven Nederlands en het Corpus Gesproken Nederlands (R5).

CONDIV-corpus geschreven Nederlands		Corpus Gesproken Nederlands (R5)	
Nederlands Nederlands (19.5 miljoen tokens)	Belgisch Nederlands (26 miljoen tokens)	Nederlands Nederlands (2 miljoen tokens)	Belgisch Nederlands (1.7 miljoen tokens)
Kwaliteitskranten (4.7 miljoen tokens)		Publieke monologen (1.1 miljoen tokens)	
Nationale. populaire kranten (5.1 miljoen tokens)		Publieke dialogen (0.7 miljoen tokens)	
Regionale. populaire kranten (7.8 miljoen tokens)		Privédialogen (1.9 miljoen tokens)	
Usenet (12.7 miljoen tokens)			
IRC (15.2 miljoen tokens)			

Langs de geografische dimensie, de horizontale as in tabel 2.1, werd een onderscheid gemaakt tussen het Nederlandse Nederlands en het Belgische Nederlands. Voor het CGN-materiaal (R5) is een specifiekere indeling mogelijk volgens verschillende streken in het Nederlandse taalgebied: de noordelijke provincies Friesland, Groningen en Drenthe; Overijssel; Gelderland; Utrecht; Noord-Holland; Zuid-Holland; Zeeland; Noord-Brabant; Nederlands Limburg; Vlaams-Brabant en Antwerpen (verder: Brabant); Oost-Vlaanderen; West-Vlaanderen; Belgisch Limburg.² Langs de stilistische dimensie kunnen verschillende lagen onderscheiden worden die we voorlopig omschrijven als een continuüm van formeel naar minder formeel taalgebruik; voor CONDIV is er enerzijds het primaire onderscheid tussen krantenmateriaal en internetmateriaal, anderzijds kan het krantenmateriaal verder opgesplitst worden in kwaliteitskranten (NRC Handelsblad en De Standaard), nationale populaire kranten (De Telegraaf en Het Laatste Nieuws) en regionale populaire kranten (De Limburger, Gazet van Antwerpen en Het Belang van Limburg) en kan het internetmateriaal verder onderverdeeld worden in Usenetmateriaal (een collectie van e-mailberichten uit nieuwsgroepen) en IRC-materiaal (Internet Relay Chat, babbelkanalen). Voor het CGN (R5) is er een opsplitsing in drie grote componenten voorzien, met name publieke monologen, publieke dialogen en privédialogen.^{3,4}

Het regionaal en stilistisch gediversifieerde materiaal dat in de beide corpora opgenomen is, zal ons in staat stellen om de contextuele dimensie (hoofdstuk 3) te testen, i.c. de invloed van de regionale afkomst en de invloed van de stilistische aspecten van het taalgebruik op de woordvolgorde na te gaan. Om louter praktische redenen zullen we ons voor de verificatie van de taalinterne dimensies (hoofdstuk 4-7) beperken tot slechts een van de corpuscomponenten, met name tot de CONDIV-corpuscomponent die het hedendaagse supraregionale materiaal van de Vlaamse kwaliteitskrant *De Standaard* bevat.⁵ Als men namelijk in aanmerking neemt dat het type woordvolgordevariatie dat hier bestudeerd wordt, frequent voorkomt (in paragraaf 2.4 zal blijken dat de volledige dataset 15287 relevante gevallen bevat), dat de meerderheid van de gegevens uit het CONDIV-corpus komen (11393 relevante observaties, d.w.z. 74.53% van alle relevante gegevens) en dat het CONDIV-corpus niet geannoteerd is, wat betekent dat de codering van de gegevens grotendeels manueel moet gebeuren, dan hoeft het geen betoog dat de codering van de volledige

dataset voor al die dimensies bijzonder tijdrovend zou zijn, overigens zonder de garantie dat de codering belangwekkende additionele resultaten zou opleveren. Om die reden is ervoor geopteerd om de taalinterne dimensies voorlopig enkel te testen op de corpuscomponent die we in het algemeen zouden kunnen omschrijven als een vertegenwoordiger van het formele Belgische Nederlands. Door de taalinterne krachten enkel te testen op krantenmateriaal van *De Standaard* zorgen we er wel voor dat de mogelijke intrusie van macroregionale (Belgisch Nederlands versus Nederlands Nederlands), stilistische (formeel versus informeel) factoren uitgesloten wordt, zodat ons onderzoek niet gecorrumpeerd wordt door de mogelijk onderliggende beïnvloeding van een (combinatie) van die taalexterne variabelen.⁶ Het voor dit onderzoek gebruikte materiaal bestaat uit artikelen die verschenen zijn in *De Standaard* in de periode van augustus tot en met december 1996.⁷ De corpuscomponent bevat in totaal iets meer dan 3,5 miljoen woordtokens (n = 3607513), hetgeen ongeveer een dertiende is van de totale omvang van het volledige CONDIV-corpus (ca. 45 miljoen woordtokens), en bestaat uit artikelen die in een van de volgende rubrieken⁸ verschenen zijn: *binnenlands nieuws* en *buitenlands nieuws*, *cultureel nieuws*, *commentaar en opinie*, *courant* en *weerpraatje*, *tv-tips* en *filatelie*.

2.3 Syntagmatische en paradigmatische restricties op het studieobject

De gegevens die voor dit onderzoek verzameld zijn, moesten aan een aantal selectievoorwaarden voldoen. Meer bepaald zijn we op zoek gegaan naar:

Tweeledige werkwoordelijke eindgroepen die opgebouwd zijn uit een deelwoord en een vervoegde vorm van het hulpwerkwoord hebben, zijn of worden voor zover die zich bevinden in een complementzin ingeleid door het grammaticaal verbindend voegwoord dat en voor zover het grammaticaal verbindend voegwoord dat niet het eerste woord van de zin is.

Door de gegevens op een zorgvuldige manier te selecteren, i.e. door een aantal selectiecriteria in te voeren die de grootte van de werkwoordelijke eindgroep, de aard van de constituerende elementen en het type bijzin controleren, hebben we trachten te vermijden dat we in een moeilijk te ontwarren complex van interagerende factoren terecht zouden komen, waarbij de impact van de factoren die in deze studie aan bod

komen, vertekend zou worden door andere, verborgen covariërende factoren (waar we voorlopig niet in geïnteresseerd zijn). Conform onze methodologische doelstelling (paragraaf 1.4) zorgen we er daarmee voor dat de interpreteerbaarheid van de resultaten vergemakkelijkt wordt en dat de repliceerbaarheid en continuïteit in toekomstig onderzoek gegarandeerd wordt—methodologische rigueur als een van de noodzakelijke voorwaarden voor inhoudelijke progressie.

Dat de set van tweeledige eindgroepen met een deelwoord en het hulpwerkwoord *zijn*, *hebben* of *worden* verkozen werd boven drie-, vier- of meerledige werkwoordelijke eindgroepen, eindgroepen met een infinitief en eindgroepen met het hulpwerkwoord *blijken*, *lijken*, *dunken*, *heten*, *voorkomen*, *behoren*, *dienen*, *moeten*, *hoeven*, *kunnen*, *mogen*, *(ge)raken*, *krijgen*, *zien*, *komen*, *blijven* (ANS 1997: 954 e.v., 1068-1069) heeft drie redenen: (i) het is een frequent bestudeerd type van variatie, waardoor we voldoende vergelijkingsmateriaal hebben (resultaten uit eerder onderzoek), (ii) het komt frequent voor in het taalgebruik en (iii) de twee mogelijke volgordes houden elkaar in voldoende mate in balans (i.t.t. bv. clusters met een infinitief en een hulpwerkwoord; vgl. De Schutter 1976: 238-241). Vanuit analytisch oogpunt is de vaststelling in (ii) en (iii) niet onbelangrijk: wil men nl. betrouwbare uitspraken doen over de mechanismen die de keuze voor de ene of de andere volgorde bepalen, dan is niet alleen de vraag naar de kwaliteit, maar ook naar de kwantiteit van de bewijslast relevant om tot betrouwbare, representatieve resultaten te komen. Daarenboven maakt een scheve verdeling in de dataset (bv. de groene [INF+AUX] versus de rode volgorde [AUX+INF] in tweeledige infinitivische clusters) de statistische modellering van de variatie technisch complexer (zie bv. Agresti 1996: 190-192).

2.4 Zoekopdracht

Met behulp van de reguliere-expressiefunctie in het corpuslinguïstische programma *Abundantia Verborum* (Speelman 1997) en een aantal speciaal voor de corpusconsultatie van het CGN (R5) door ons ontwikkelde Pythonscripts⁹ (versie 2.2.2, april 2003) werden de relevante attestaties uit de corpora geëxtraheerd (zie appendix B voor de technische gegevens). Om de invloed van de context optimaal te kunnen bestuderen, werden niet alleen de targetzinnen, maar ook de 10 voorafgaande

en 10 volgende paragrafen van elke targetzin meegenomen in de dataset (enkel voor CONDIV; het CGN-materiaal (R5) zal niet gebruikt worden voor een discursieve analyse).¹⁰ De zoekopdracht leverde voor het CONDIV-corpus 20653 observaties op, voor CGN (R5) 21192 attestaties, wat in een totaal van 41845 observaties resulteert. Voor de subcomponent *De Standaard hedendaags, supraregionaal* leverde de zoekopdracht 4719 observaties op. Van al die observaties kan gezegd worden dat ze formeel aan de eisen van onze zoekopdracht voldoen. Het betekent echter niet dat alle observaties vanuit linguïstisch oogpunt relevant zijn, zodat een zuiveringsoperatie noodzakelijk was. In de volgende paragraaf stellen we een beknopte typologie van irrelevante observaties op.

2.5 Zuivering van de dataset

Dat de patronen die de zoekmachine op basis van de zoekopdracht uit het CONDIV-corpus gehaald heeft niet allemaal relevant zijn ten aanzien van de initieel opgestelde (linguïstische) selectiecriteria, heeft vier oorzaken: (i) de linguïstisch geformuleerde selectiecriteria doen een beroep op formeel onherkenbare semantische en/of syntactisch-categoriale kenmerken (bv. het grammaticaal verbindend voegwoord *dat* onderscheidt zich formeel niet van het relativum *dat*); (ii) de formele afbakening van het deelwoord en hulpwerkwoord was wel algemeen genoeg, maar niet distinctief genoeg (bv. een meervoudig nomen als *gesteenten* kon op basis van de door ons geformuleerde zoekopdracht niet geweerd worden); (iii) met de structurele posities waar geen inhoudelijk-linguïstische beperkingen opgelegd worden, het middenstuk en de laatste zinsplaats, wordt de deur opengezet voor de introductie van ongewenste elementen die de aard van de constructie dermate veranderen dat ze irrelevant wordt (bv. als het laatste woord van het middenstuk of het eerste woord van de laatste zinsplaats een werkwoord is, dan hebben we niet meer te maken met een tweeledige, maar met een drieledige verbale cluster); (iv) de aard en de kwaliteit van het materiaal in het corpus zorgt er onafhankelijk van de overgeneraliserende tendens van de zoekopdracht voor dat een aantal observaties irrelevant is. Een groot deel van de irrelevante observaties kon vrij makkelijk gedetecteerd worden (ANS 1997 werd gebruikt bij twijfel). Het gaat dan meer bepaald over oneigenlijke hulpwerkwoorden (bv. *zijn* als pronomen) en oneigenlijke deelwoorden (substantieven, bijwoorden,

determinatoren en conjuncties), het voegwoord *dat* dat als aanwijzend voornaamwoord, betrekkelijk voornaamwoord, redengevend voegwoord of graadaanduidend voegwoord gebruikt is, eindgroepen met infinitieven, drie- en vierledige eindgroepen, eindgroepen die zich in een hoofdzin of in andere types bijzinnen bevinden, (oneigenlijke) eindgroepen waarvan de constituerende delen tot verschillende zinnen behoren en attestaties die wellicht door veelvuldige bewerking een ongrammaticale ellips of een ongrammaticale opeenhoping van werkwoorden bevatten. Voor een ander deel van de irrelevante attestaties was het minder makkelijk uit te maken of ze als relevant beschouwd moesten worden. We stellen een beknopte typologie op van problematische gevallen en we voegen er de criteria bij die gebruikt zijn om de relevantie voor het huidige onderzoek te bepalen:

- Het grammaticaal verbindend voegwoord *dat* kan deel uitmaken van een voegwoordelijke uitdrukking (bv. *op voorwaarde dat*; ANS 1997: 543). Om ervoor te zorgen dat de complexe conjuncties geen onderliggende invloed zouden uitoefenen op de resultaten in de volgende hoofdstukken, werd ervoor geopteerd om ze uit te sluiten. Daarvoor werd de lijst in ANS (1997: 1170-1175) en de CONST-lijst van Smessaert gebruikt (p.c.; zie ook Beeken & Smessaert 1996), waarin 97 (voor dit onderzoek relevante) voegwoordelijke uitdrukkingen opgenomen zijn. Een speciale voegwoordconstructie is de zgn. *feit dat*-constructie:

- (5) De effectieve consulaire hulp dankte hij aan het feit dat hij in Casablanca werd veroordeeld en opgesloten. (DS)

Volgens Smessaert (p.c.) zijn sommige *feit dat*-constructies als voegwoordconstructies te beschouwen (vermoedelijk als het gevolg van een grammaticalisatieproces), andere niet. Criteria zijn echter niet zo gauw te geven en daarom werd besloten om alle *feit dat*-constructies voorlopig als voegwoordelijke constructies te beschouwen en dus uit te sluiten uit de dataset.

- Deelwoorden kunnen in bepaalde contexten adjectivische karakteristieken hebben. Er is voor geopteerd om enkel de adjectivische deelwoorden uit de dataset te verwijderen die gemodificeerd worden door een typisch adjectivisch vrij of gebonden morfeem of reeks van morfemen. De volgende types adjectieven werden daarom uit onze dataset verwijderd: adjectieven in de vorm

van een tegenwoordig deelwoord (bv. *behoudend*), adjectieven in de vorm van een deelwoord met een typisch adjectivisch prefix of in verbogen vorm (bv. *onaangetast*, *geplande*), adjectieven in de vorm van een deelwoord dat ingeleid wordt door een of meerdere adjectiefmodicerende bijwoorden (bv. *te geïnteresseerd*), complex samengestelde adjectieven waarvan een deelwoord of een op een deelwoord gelijkend adjectief een van de leden is en waarbij het geheel niet tot een werkwoordelijk paradigma behoort (bv. *arbeidsongeschikt*). Dat betekent bijvoorbeeld dat onbetwistbaar adjectivische deelwoorden zoals *bestand* en *bereid* aanvaard worden in de dataset, althans voor zover ze niet gemodificeerd werden door een van de bovenvermelde morfemen.

- Werkwoordelijke eindgroepen met twee deelwoorden en een hulpwerkwoord kunnen zowel twee- als driedig beschouwd worden (bv. *gedragen zijn geweest*). In het eerste geval wordt een van de deelwoorden als een adjectief opgevat, in het tweede geval zijn alle deelwoorden werkwoordelijk. Omdat niet eenvoudig uit te maken is wat een deelwoord tot adjectief maakt (zie ook hierboven), worden dergelijke constructies niet aanvaard in de dataset, althans voor zover de beide deelwoorden deel uitmaken van een werkwoordelijk paradigma—en dus terugvoerbaar zijn op een infinitief.
- Het CONDIV-corpus bleek soms meer dan een keer dezelfde zin te bevatten. De reden daarvoor is vermoedelijk te zoeken in het feit dat een krant soms meer dan een artikel over hetzelfde onderwerp bevat; zo is het bijvoorbeeld goed mogelijk dat een krant een bepaald onderwerp op de voorpagina introduceert en er verderop in de krant dieper op ingaat en dat de redacteur/redacteurs daarbij bepaalde stukken kopieert/kopiëren. Ze werden uit de dataset verwijderd om geen vertekening van het algemene beeld te veroorzaken.

De gegevens die uit het CGN (R5) gehaald werden, voldeden ondanks de woordsoortannotatie ook niet allemaal aan de selectiecriteria, doordat de zoekopdracht veel minder gespecificeerd was dan de CONDIV-zoekopdracht (zie appendix B). De volgende types irrelevante observaties werden precies om die reden (grotendeels automatisch) uit de dataset verwijderd:

- De verbale cluster bevindt zich in een hoofdzin.

- Er bevinden zich andere types hulpwerkwoorden in de verbale cluster.
- De bijzin begint niet met het grammaticaal verbindend voegwoord *dat* (zie hierboven).
- De verbale cluster is drie- of vierledig.

Specifiek aan het CGN (R5) waren irrelevante observaties waarvoor na een volledige manuele controle vastgesteld kon worden dat ze fout geannoteerd waren: *gezien* als conjunctie in plaats van deelwoord, *gegeven* als adjectief in plaats van deelwoord (zie hierboven), *dat* als betrekkelijk voornaamwoord in plaats van als onderschikkend voegwoord, etc. Verder kwamen er in de CGN-dataset (R5) een aantal typische spreektaalfenomenen voor die voor de irrelevantie van de observatie zorgden: het ging daarbij om gevallen met expletief *dat*, het gebruik van een hoofdzin waar een bijzin verwacht wordt (*dat iedereen al dacht de Ronde is gedaan*), een doorbroken ‘complex’ voegwoord (*na euh dat*) en een ontbrekende matrixzin.

Na de zuiveringsoperatie bleven voor het CONDIV-corpus 11393 relevante observaties over, 3894 relevante attestaties voor het CGN (R5), wat in een totaal van 15287 observaties resulteert. De dataset met enkel de gegevens uit de subcomponent *De Standaard supraregionaal* bevat 2390 relevante observaties.

2.6 Codering van de attestaties en statistische verwerking

Na het verwijderen van de irrelevante observaties uit de dataset was alles klaar voor de codering van de zinnen. De aard van de factoren bepaalde of een volautomatische of semi-automatische codering (op basis van formele kenmerken) uitgevoerd kon worden, dan wel of alle attestaties manueel gecodeerd dienden te worden. De codering van de attestaties vond plaats met behulp van het labelmechanisme in *Abundantia Verborum*. Per factor werd daarbij aan elke zin één waarde toegekend: voor de factor *regio* is dat bijvoorbeeld *Belgisch Nederlands* of *Nederlands Nederlands*; voor sommige andere factoren lag de keuze van het aantal en de aard van de waarden minder voor de hand (de problemen met de operationaliseringskwesaties behandelen we niet hier, maar komen in de betreffende hoofdstukken aan bod). Op die manier krijgen we een datamatrix met voor elke zin een combinatie van waarden (het aantal waarden is gelijk aan het aantal factoren).¹¹

Om die waardenreeksen samen te vatten, grafisch te representeren en om na te gaan of er belangwekkende associaties en interacties zijn tussen bepaalde variabelen (over alle zinnen heen), zullen we gebruik maken van een hele reeks descriptieve en inferentiële statistische technieken die beschikbaar zijn in het statistische softwarepakket R (2003), versie 2.0.1 (zie de analytische doelstelling in paragraaf 1.4). De nadruk zal daarbij steeds op de interpretatie van het resultaat liggen en niet op de technische details (zie daarvoor bv. Agresti 1996 en Harrell 2001). Voor alle statistische toetsen die verderop gebruikt zullen worden, is het belangrijk om te onthouden dat het gebruik van statistische tests geen doel op zich is, maar wel een middel om op basis van een steekproef betrouwbare uitspraken te doen over de volledige populatie van tweeledige werkwoordelijke eindgroepen die aan de genoemde selectiecriteria voldoen.¹² Dat laatste betekent bijvoorbeeld dat we op basis van onze taalexterne resultaten in hoofdstuk 3 enkel uitspraken zullen kunnen doen over het contemporaine Nederlands zoals dat in kranten, op het internet en in gesproken vorm gebruikt wordt. Hoe de situatie er in bepaalde types literair Nederlands uitziet, om maar een voorbeeld te noemen, is een vraag die hier niet beantwoord zal kunnen worden. Voor de taalinterne dimensies is de interpretatieradius nog kleiner, de resultaten zullen nl. enkel betrekking hebben op het hedendaagse formele Belgische Nederlands.

De verschillende statistische tests die we vanaf hoofdstuk 3 tot en met hoofdstuk 7 zullen gebruiken, worden voorgesteld en besproken wanneer ze voor het eerst gebruikt worden. Concreet komt het erop neer dat heel wat van de categorische bivariate en conditionele (trivariate) tests voor het eerst uitgelegd worden in hoofdstuk 3, tests voor de analyse van continue gegevens komen voor het eerst ter sprake in hoofdstuk 4. In hoofdstuk 7 maken we dan ten slotte de sprong naar de statistische modellering van de variatie, i.e. zal het verband beschreven worden tussen de rode [AUX+PART] of de groene volgorde [PART+AUX] en een reeks factoren (multivariate tests). De tekst is daarbij zo opgesteld dat hij niet enkel interessant is voor onderzoekers die reeds enige ervaring hebben met de statistische behandeling van een dergelijke materie, maar ook voor diegenen die niet zo vertrouwd zijn met statistische bewerkingen: de toegepaste berekeningswijze wordt telkens toegelicht op een wijze die geen voorkennis veronderstelt. Om de geïnteresseerde lezer de kans te bieden de in dit proefschrift uitgevoerde statistische analyses over te doen, stellen we

de datamatrices die aan dit onderzoek ten grondslag liggen ter beschikking via <http://www.ling.arts.kuleuven.ac.be/qlvl/datasets>. Ten behoeve van de lezer die, zonder zich te verdiepen in de methodologische werkwijze, enkel een overzicht wil van de belangrijkste linguïstische resultaten, hebben we vanaf hoofdstuk 3 tot en met hoofdstuk 6 aan het eind van elk hoofdstuk een samenvatting toegevoegd met een overzicht van de belangrijkste ‘linguïstische’ vondsten.

Noten:

1. Het CONDIV-corpus bevat ook nog een diachrone dimensie, doordat er naast materiaal uit de jaren 1990 ook materiaal uit de jaren 1950 en 1970 opgenomen is. Aangezien we met deze studie een synchrone analyse van de woordvolgordevariatie willen bieden, laten we de diachrone dimensie buiten beschouwing en concentreren we ons enkel op het 1990-materiaal van het CONDIV-corpus (zie Grondelaers et al. 2000 voor een volledige corpusbeschrijving).
2. Het CGN (R5) operationaliseert de regionale afkomst van de sprekers op drie verschillende manieren, namelijk de regio waar de taalgebruiker geboren is, de regio waar de taalgebruiker opgegroeid is en de regio waar de taalgebruiker op het ogenblik van de dataopname woonachtig is. Vooraf werd voor elke regio apart de nulhypothese getest dat de verdeling van de attestaties over de groene [PART+AUX] en de rode volgorde [AUX+PART] voor elk van de drie onderscheiden regiovariabelen stabiel blijft (m.b.v. een chi-kwadraatanalyse). Zo werd bijvoorbeeld voor Noord-Holland nagegaan of er statistisch significante verschillen waren in het gebruik van de groene [PART+AUX] en de rode volgorde [AUX+PART] tussen de regio waar men geboren is, waar men opgegroeid is en waar men op het moment van de dataopname resideerde. Voor alle regio's werd de nulhypothese niet verworpen ($p > .05$), hetgeen betekent dat er geen statistisch significante verschillen zijn tussen de effecten van de drie verschillende types regio. Aangezien we die regionale variabelen slechts in één specifieke bivariate analyse willen gebruiken (zie verderop), heeft het geen zin om de drie variabelen te selecteren. Daarom gebruiken we in wat volgt slechts een van de drie regionale variabelen, m.n. de plaats waar de taalgebruiker opgegroeid is. In dat verband moet ook nog opgemerkt worden dat het eigenlijk ook voor een deel van het CONDIV-corpus, m.n. het IRC-materiaal, mogelijk is om een specifiek- regionale hypothese te testen, aangezien er chatkanalen bestaan als #Leuven, #Amsterdam en dergelijke. Echter, in dat materiaal kan in tegenstelling tot de CGN-gegevens (R5) niet gegarandeerd worden dat de taalgebruikers daadwerkelijk uit Leuven of Amsterdam komen, en daarom wordt van een verdere regionale opsplitsing van het IRC-materiaal afgezien.
3. Het CGN (R5) is langs de stilistische dimensie samengesteld uit 14 componenten. Wij hebben, zoals het CGN (R5) zelf ook doet, een aantal van die componenten samengenomen in grotere gehelen, enerzijds omdat het aantal beschikbare gegevens per component soms nogal verschilt tussen Vlaanderen en Nederland, anderzijds omdat sommige verschillen tussen afzonderlijke

componenten niet relevant zijn voor dit onderzoek. Opgemerkt moet nog worden dat in de vijfde release de component privémonologen niet beschikbaar is.

4. De afkortingen die in het vervolg van dit proefschrift gebruikt zullen worden om de verschillende corpuscomponenten aan te duiden zijn: HLN voor Het Laatste Nieuws, DS voor De Standaard, HBVL voor Het Belang van Limburg, GVA voor Gazet van Antwerpen, NRC voor NRC Handelsblad, DT voor De Telegraaf, DL voor De Limburger, Use-BL voor Usenet-België, Use-NL voor Usenet-Nederland, IRC-BL voor IRC-België, IRC-NL voor IRC-Nederland; *priv_dia* voor privédialogen, *pub_dia* voor publieke dialogen en *pub_mon* voor publieke monologen.
5. Niet het hele corpussegment met het *De Standaard*-materiaal wordt voor dit onderzoek gebruikt, maar enkel het hedendaagse (artikelen van 1996), supraregionale materiaal. Daarnaast bevat het corpussegment met artikelen van *De Standaard* ook nog artikelen uit 1958 en 1978 (diachrone dimensie) en technische artikelen over o.a. economische onderwerpen en sportartikelen (intern-stilistische dimensie). Het 1958- en 1978-materiaal werd niet voor dit onderzoek gebruikt, om de voor de hand liggende reden dat het onderzoek een uitsluitend synchroon-hedendaags perspectief inneemt; de technische artikelen werden wegens de specificiteit van het taalgebruik uit het onderzoek geweerd.
6. Uiteraard is daarmee niet gezegd dat er absoluut geen sprake meer van kan zijn dat een macro- of microregionale of stilistische factor werkzaam is in de dataset die we zullen onttrekken aan de corpuscomponent *De Standaard*. Zo kan bijvoorbeeld de eventuele invloed van de regionale afkomst van de verschillende journalisten niet gecontroleerd worden. Wij gaan ervan uit dat een volledige controle van de dataset in wezen onmogelijk is en slechts nuttig is indien er duidelijk aanwijzingen zijn om een extra controle uit te voeren. Op basis van de beschikbare literatuur kunnen dergelijke aanwijzingen niet gevonden worden.
7. In feite bevat het officiële CONDIV-corporus enkel *De Standaard*-materiaal van de maanden november en december 1996. Voor de studie naar de taalstructurele effecten is dat materiaal echter uitgebreid met de maanden augustus tot en met oktober. De selectieprocedures en opmaak van die uitbreiding verschilt in geen enkel opzicht van de oorspronkelijke *De Standaard*-component van het CONDIV-corporus.
8. De onderverdeling van de artikelen in de verschillende rubrieken gebeurde op de redactie van *De Standaard* en werd door de samenstellers van het CONDIV-corporus overgenomen.
9. Python is, net als PERL, een scriptingtaal die in corpuslinguïstische kringen vaak gebruikt wordt om ad hoc problemen op te lossen (daar waar *Abundantia Verborum* geen oplossing bood). Het spreekt voor zich dat de beide scriptingtalen daarnaast ook nog andere toepassingen hebben: <http://www.perl.com/> en <http://www.python.org/>.
10. De selectie van de 10 voorafgaande en 10 volgende paragrafen was wel aan de voorwaarde gebonden dat de paragrafen tot dezelfde discursieve eenheid (bv. tot hetzelfde krantenartikel) als de targetzin dienden te behoren. Bij de bestudering van de context is het nl. onaanvaardbaar dat de keuze van de volgorde afhankelijk gemaakt zou worden van een andere discursieve eenheid (die

door iemand anders, op een andere plek en op een ander tijdstip geschreven is). Met een PERL-script werden de discursieve grenzen geïdentificeerd en werd ervoor gezorgd dat de voorafgaande en volgende context nooit een dergelijke grens kon overschrijden.

11. De datasets zijn beschikbaar via <http://www.ling.arts.kuleuven.ac.be/qlvl/datasets>.
12. Als verderop gesproken wordt over de volledige populatie van alle woordvolgordes in werkwoordelijke eindgroepen, dan is daarmee enkel de subset bedoeld die voldoet aan de hierboven omschreven selectiecriteria.

Hoofdstuk 3

De contextuele dimensie

3.1 Inleiding

We starten ons onderzoek naar de factoren die de keuze voor de groene [PART+AUX] of rode [AUX+PART] volgorde bepalen met een van de meest opvallende en best bestudeerde dimensies in het woordvolgordeonderzoek, m.n. de contextuele dimensie. Uit het overzicht van de algemene (1.3.1) en taalwetenschappelijke literatuur (1.3.4) is al duidelijk naar voren gekomen dat de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde in het verleden al vaker in verband gebracht is met sociostilistische en regionale factoren. Illustratief daarvoor is het volgende citaat uit de ANS (1997: 1067-1069): “Allereerst tekent er zich een verschil tussen gesproken en geschreven taal af. De gebruikelijkste volgorde in gesproken taal, zeker in minder formeel taalgebruik, is die met het deelwoord eerst [...], al is de mate waarin die volgorde gebruikt wordt niet in alle delen van het taalgebied dezelfde. In geschreven taal bestaat er een tendens om vaker de omgekeerde volgorde te gebruiken [...]. Die tendens is het duidelijkst in journalistieke teksten”.

In dit hoofdstuk willen we nagaan in hoeverre de resultaten van het vroegere onderzoek bevestigd kunnen worden aan de hand van onze steekproef en willen we een aanvulling annex verfijning bieden van de kennis die er op dit moment bestaat over de regionale en stilistische effecten op de woordvolgorde in de tweeledige werkwoordelijke eindgroep. Om alvast een eerste idee te geven in welke richting dit hoofdstuk zal evolueren, wordt in tabel 3.1 een overzicht geboden van de factoren die in dit hoofdstuk aan bod zullen komen:

Tabel 3.1: Overzicht van de factoren in dit hoofdstuk.

Regio	Register
Staat	Modus
Provincie	Mate van interactie
	Mate van redactionele controle

Het materiaal in dit hoofdstuk is als volgt over de verschillende paragrafen verdeeld: in paragraaf 3.2 zal de impact van regionale verschillen op de keuze van woordvolgorde bestudeerd worden, in paragraaf 3.3 komt de invloed van het register aan bod. In elk van die paragrafen wordt eerst stilgestaan bij de stand van zaken in de wetenschappelijke literatuur en de hypothesen die daaruit voortvloeien. Vervolgens wordt het empirische onderzoek uitgevoerd. In paragraaf 3.4 wordt het verband tussen de beide effecten bestudeerd en in paragraaf 3.5 wordt een uitgebreide samenvatting van dit hoofdstuk gegeven.

Vooraleer het onderzoek naar de contextuele beïnvloeding aan te vatten, staan we eerst nog even stil bij de steekproef die we voor het onderzoek gaan gebruiken. Zoals in hoofdstuk 2 aangekondigd werd, verschilt de empirische basis in dit hoofdstuk grondig van die in de volgende, taalinterne hoofdstukken: om de invloed van regio en register ten volle te kunnen bestuderen, werd er namelijk voor geopteerd om het volledige CONDIV-corpus en de vijfde release van het Corpus Gesproken Nederlands (althans het gedeelte dat met woordsoortinformatie verrijkt is) te gebruiken. De zoekopdracht in die beide corpora, rekening houdend met de gestelde selectiecriteria, leverde 15287 observaties op. In tabel 3.2 wordt de verdeling van de data over de twee syntactische varianten weergegeven. Daaruit blijkt dat de groene [PART+AUX] en rode [AUX+PART] verbale clusters in het materiaal respectievelijk ongeveer een derde en twee derde van de dataset in beslag nemen.

Tabel 3.2: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde in de dataset.

Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
37.3%	62.7%
(5709/15287)	(9578/15287)

Om na te gaan in hoeverre de steekproefproporties (i.e. de proporties van de beide syntactische alternatieven in onze steekproef van 15287 observaties) in tabel 3.2 representatief zijn voor de (onbekende) proporties in de volledige populatie (i.e. de proportie van *alle* werkwoordelijke eindgroepen die aan de vooropgestelde selectiecriteria voldoen), bakenen we 95%-betrouwbaarheidsintervallen (verder: B.I.) af rond de twee steekproefproporties. Die B.I. reiken van 36.6% tot 38.1% voor de proportie groene woordvolgordes [PART+AUX] en van 61.9% tot 63.4% voor de proportie rode volgordes [AUX+PART]. Dat betekent dat we met 95% zekerheid kunnen zeggen dat de proportie van groene woordvolgordes [PART+AUX] in de hele populatie ergens tussen 36.6% en 38.1% zal liggen en de populatieproportie rode volgordes [AUX+PART] tussen 61.9% en 63.4%.¹ Daaruit kunnen we besluiten dat de steekproefproporties in tabel 3.2 behoorlijk representatief zijn voor de populatieproporties en dat het grote verschil tussen de rode [AUX+PART] en de groene [PART+AUX] volgorde een reëel verschil is. Dat verschil is overigens statistisch significant ($\chi^2 = 979.21$, $df = 1$, $p < .0001$), wat betekent dat de kans dat een dergelijk extreem verschil te wijten is aan het toeval kleiner is dan 1 op 10000. Met 99.99% zekerheid kunnen we dan zeggen dat de rode volgorde [AUX+PART] vaker voorkomt dan de groene woordvolgorde [PART+AUX]. Dat laatste moeten we natuurlijk interpreteren tegen de achtergrond van het type data dat voor deze studie gebruikt is: een groot deel van onze dataset bestaat nl. uit geschreven krantenmateriaal en (voorbereide) gesproken monologen. Uit het vervolg van dit hoofdstuk zal blijken dat het precies die bias ten voordele van het formelere taalgebruik is die de dominantie van de rode volgorde [AUX+PART] in de hand werkt. In paragraaf 3.3 gaan we bij de bespreking van de resultaten van de registervariabele dieper in op die kwestie.

3.2 De impact van regio

In deze sectie worden de resultaten van ons corpusonderzoek naar de impact van regio op de keuze voor de groene [PART+AUX] of rode [AUX+PART] volgorde gepresenteerd. Eerst bespreken we de belangrijkste literatuur ter zake (paragraaf 3.2.1), vervolgens worden de verschillende hypothesen voorgesteld (paragraaf 3.2.2) en tot slot presenteren en evalueren we de resultaten van ons onderzoek (paragraaf 3.2.3), dat enerzijds naar de preferentieverschillen tussen het Belgische en het Nederlandse Nederlands peilt (paragraaf 3.2.3.1) en anderzijds naar de preferentieverschillen tussen de verschillende (dialectologische) streken binnen het Nederlandse taalgebied (paragraaf 3.2.3.2).

3.2.1 *Status quaestionis*

Ondanks de voorliefde van de dialectologie voor fonetische en lexicale variatie is voor het Nederlandse taalgebied toch heel wat geweten over de geografische verspreiding van syntactische varianten, waaronder de variatie in de werkwoordelijke eindgroep. Voor de studie van regionale verschillen is zowel een microperspectivische als een macroperspectivische aanpak gehanteerd, i.e. respectievelijk de studie van regionale verschillen op het niveau van de individuele dialecten en de studie van regionale verschillen op het niveau van de standaardtaal, c.q. het verschil tussen de supraregionale variëteit die in Vlaanderen (verder: Belgisch Nederlands) gebezigd wordt en die in Nederland (verder: Nederlands Nederlands).

Dat *microperspectivische onderzoek*, dat vooral een aanvang genomen heeft sinds de beroemde dialectologische studie van Pauwels (1953), heeft een aantal dingen duidelijk gemaakt:²

- De groene volgorde [PART+AUX] is frequenter dan de rode volgorde [AUX+PART] (een aantal uitzonderingen niet te na gesproken; zie de literatuur in noot 2 voor een opsomming van de ‘rode’ gebieden in het taalgebied).
- De groene volgorde [PART+AUX] is vooral populair in de noordelijke provincies van Nederland (Friesland, Groningen, Drenthe). Over de andere gebieden lopen de meningen en resultaten in min of meerdere mate uiteen van

een extreme voorkeur voor de groene volgorde [PART+AUX] (Frings & Vandenheuvel 1921 voor Vlaanderen, Stroop 1970 voor Nederland) tot een meer gematigde voorkeur voor de groene volgorde [PART+AUX] (Pauwels 1953, Gerritsen 1991). De redenen voor de verschillen kunnen vermoedelijk gevonden worden in de verschillende methodes en de wijze van dataselectie (zie Stroop 1970 voor een kritiek op Pauwels 1953; zie Gerritsen 1991 voor een verklaring van de verschillen met Stroop 1970).

- In vergelijking met de andere gebieden valt het op dat de rode volgorde [AUX+PART] erg frequent gebruikt wordt in de Nederlands-Limburgse dialecten (i.t.t. wat Renkema 1995 en Lamers 1997 beweren; paragraaf 1.3.1.2). Dat is des te opvallender, omdat het Nederlands-Limburgse dialectgebied onmiddellijk aansluit bij het Duitse taalgebied, waar, althans in het Hoogduits, enkel de groene woordvolgorde [PART+AUX] gebruikt wordt. Dat verschil, dat lijkt samen te lopen met de rijksgrens tussen Nederland en Duitsland, kan op twee manieren geïnterpreteerd worden: (1) de Limburgers trachten zich af te zetten tegen de burens van over de grens en ze maken daarbij gebruik van een variant die wel in het Standaardnederlands mogelijk is, maar niet in het Hoogduits (vgl. de opvatting in een aantal oudere taaladvieswerken dat de groene volgorde [PART+AUX] een germanisme zou zijn); (2) door het gebrek aan syntactische dialectstudies van de Ripuarische of Duits-Nederfrankische dialecten hebben Duitse dialectologen te weinig zicht op mogelijke variatie in de positionering van de werkwoordelijke elementen in de bijzin. Dat de tweede interpretatie niet onaannemelijk is, blijkt al na raadpleging van een aantal studies dat zich onder meer op de syntaxis van een bepaald Duits dialect concentreert: zo is op een van de kaarten in de *Sprachatlas der deutschen Schweiz* (1975, III.261) te zien dat in het zuidwesten van Zwitserland vaker de rode [AUX+PART] dan de groene [PART+AUX] volgorde gebruikt wordt (*bin gewesen*; zie ook Hodler 1969). Ook in het Nürnbergs (Gebhardt 1907), het Mönchengladbachs (Walter & Gietzen 2003), in Oost-Thüringen (Sperschneider 1959) en in de Zwabische dialecten van Zuid-Hongarije (Wild 1997) blijkt de rode volgorde [AUX+PART] al eens voor te komen.

Het *macroperspectivische onderzoek*, dat de regionale verspreiding van de beide syntactische varianten in de bovenregionale variëteiten van Vlaanderen en Nederland als onderzoeksobject heeft genomen, levert tegenstrijdige resultaten op: Verhasselt (1961) meent dat Vlaanderen minder vaak de rode volgorde [AUX+PART] gebruikt dan Nederland, een tendens die Stroobants (1997: 147) geverifieerd heeft voor geschreven reclamemateriaal. Op basis van de cijfers die De Schutter (1996: 213) geeft, zien we dezelfde tendens ook in het formele gesproken Nederlands van televisie-interviews en -gesprekken. In tegenstelling tot de vorige drie studies komt Haeseryn (1990: 157) op basis van vier gedragsexperimenten echter tot de conclusie dat er geen sprake kan zijn van “een systematische (en significante) tegenstelling tussen Vlaanderen als geheel en Nederland als geheel [...]”, al constateert hij wel dat de West-Vlamingen een opvallend sterke neiging hebben om de groene volgorde [PART+AUX] te gebruiken.

De reden waarom Haeseryn geen significant verschil tussen Vlaanderen en Nederland vindt, is niet eenvoudig uit te maken. Wat wel duidelijk is, is dat zijn resultaten in zekere mate gecorrumpeerd zouden kunnen zijn door wat in de psycholinguïstiek bekend staat als een zgn. *syntactic priming* effect (Bock 1986; ook bekend als *structural priming* of *syntactic persistence*: zie Pickering et al 2000: 205-206). Daarmee bedoelt men dat taalgebruikers de neiging hebben om een bepaald syntactisch patroon dat recent geactiveerd is, hetzij actief, hetzij passief, te hergebruiken.³ Voor het onderzoek van Haeseryn (1990) betekent dat dat de proefpersonen zich bij de productie van werkwoordelijke eindgroepen mogelijkwijs hebben laten leiden door wat hen in de opgave aangeboden werd. Voor twee van de vier tests, de zgn. *compliance tests*, is Haeseryn (1990) zich daar ook van bewust. In die tests, waarin de testpersonen gevraagd werd om zinnen vragend en ontkennend te maken, wordt nl. de mate van meegaandheid van de testpersonen gemeten, i.e. in welke mate veranderen de testpersonen, naast de noodzakelijke wijzigingen om een zin vragend of ontkennend te maken, ook de aangeboden volgorde van de werkwoordelijke elementen in de eindgroep?⁴ In tabel 3.3 vatten we de resultaten van de mate van meegaandheid voor de *compliance test* items samen (enkel de items die een tweeledige werkwoordelijke eindgroep met een deelwoord en een hulpwerkwoord bevatten: items A106, A305, B104, B106, B113 en B306; gebaseerd op tabel 4.3 in Haeseryn (1990: 143)):

Tabel 3.3: Resultaten van de mate van meegaandheid in Haeseryn (1990).

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Aangeboden: [PART+AUX]	89.78 % (703/783)	10.22 % (80/783)
Aangeboden: [AUX+PART]	29.18 % (227/778)	70.82 % (551/778)

Uit tabel 3.3 blijkt duidelijk dat de volgorde van het hulpwerkwoord en het deelwoord in het merendeel van de aangeboden zinnen niet veranderd werd en dat met andere woorden de mate van meegaandheid hoog is: als een groene woordvolgorde [PART+AUX] aangeboden werd, werd in bijna 90% van de gevallen een groene woordvolgorde [PART+AUX] overgenomen; als een rode woordvolgorde [AUX+PART] aangeboden werd, werd in 70% van de gevallen een rode woordvolgorde [AUX+PART] overgenomen (de mate van meegaandheid verschilt dus wel naargelang van welke volgorde aangeboden werd, i.e. de rode volgorde [AUX+PART] werd vaker veranderd in een groene volgorde [PART+AUX] dan de groene volgorde [PART+AUX] in een rode volgorde [AUX+PART]).

In tegenstelling tot het eerste type test veronderstelt Haeseryn van het tweede type test, de zgn. *selection test* (i.c. een onderschikkings- en puzzeltest), dat de volgorde die de testpersonen kiezen niet beïnvloed wordt door de aangeboden volgorde (Haeseryn 1990: 142), omdat er steeds voor gezorgd is dat de verbale elementen niet onmiddellijk in elkaars buurt staan. Toch zijn wij van mening dat ook in een van de twee tests, met name in de puzzeltest, sprake geweest zou kunnen zijn van syntactisch kopieergedrag. In de puzzeltest, waarbij de onderdelen van een zin door elkaar gehaspeld worden en waarbij de informanten nadien gevraagd wordt de delen weer in een voor het Nederlands aanvaardbare volgorde te plaatsen, moet de onderzoeker nl. de keuze maken in welke volgorde de verschillende onderdelen aangeboden worden. Haeseryn (1990) heeft daarbij consequent gekozen voor een scheiding van het hulpwerkwoord en het deelwoord door ander materiaal, maar dat neemt niet weg dat dan nog steeds een van die verbale elementen als eerste aangeboden moet worden. Eigen onderzoek (paragraaf 5.2.4) heeft duidelijk gemaakt dat informanten zich in dergelijke omstandigheden laten leiden door wat eerst

aangeboden wordt: staat het deelwoord eerst, dan wordt significant vaker de groene volgorde [PART+AUX] gekozen, staat het hulpwerkwoord eerst, dan stijgt de kans op een rode volgorde significant. Een en ander heeft te maken met een nog gebrekkige kennis over hoe taalgebruikers een dergelijke test precies uitvoeren, i.c. over het mechanisme van het puzzelen in een experimentele omgeving.

Een andere reden waarom de macroregionale resultaten van Haeseryn (1990) afwijken van de andere onderzoeken kan uiteraard ook in het experimentele karakter van zijn studie liggen: in hoeverre verschilt het reële taalgebruik van dat dat in experimentele omstandigheden tot stand gekomen is (dat suggereert Haeseryn 1990: 346-349 ook zelf)? Tot slot kan er ook een leereffect opgetreden zijn, waarmee bedoeld wordt dat het niet uit te sluiten is dat de resultaten van het onderzoek beïnvloed zijn door het groeiende bewustzijn van de informant over het studieobject. Het gevolg daarvan kan zijn dat het experiment veeleer naar taalnormen peilt dan naar taalgedrag. De grote moeilijkheid van experimenteel onderzoek is namelijk een onbevangen informant onbevangen te houden. Psycholinguïsten doen dat door een veelvoud van afleiders at random te distribueren over de experimentele items, zodat het onderzoeksobject ‘verdrinkt’ in de andere items en de informant niet kan achterhalen wat het echte onderzoeksobject is. Haeseryn (1990) doet dat slechts in beperkte mate, wellicht vanuit de praktische noodzaak om het aantal testitems binnen de perken te houden, gecombineerd met de doelstelling om naar allerlei aspecten van de werkwoordsvolgorde te peilen (tweeledige eindgroepen met infinitieven en deelwoorden, drie- en meerledige eindgroepen, doorbroken eindgroepen en splitsing van scheidbaar samengestelde werkwoorden). Dat heeft als gevolg dat een heel aantal variabelen over een beperkt aantal testitems verdeeld moet worden. Vanuit dat opzicht is het o.i. onbegrijpelijk dat Haeseryn de bijzonder grote pool van informanten (526 informanten!) slechts onderverdeeld heeft in twee groepen. Indien hij meerdere groepen had onderscheiden die hij elk een deel van de testitems had kunnen geven, dan had hij niet alleen meer testitems per variabele kunnen opnemen in het design, maar dan had hij vooral meer afleiders in het design kunnen stoppen. Uiteraard moet toekomstig onderzoek uitwijzen in hoeverre de hier geformuleerde bedenkingen terecht zijn. Het moet echter duidelijk zijn dat de resultaten van het onderzoek van Haeseryn (1990) in het vervolg van dit proefschrift tegen de achtergrond van de hier gemaakte opmerkingen geïnterpreteerd moeten worden.

3.2.2 Hypotheses

De volgende hypothesen, geëxtraheerd uit de resultaten van het vorige onderzoek, zullen aan de grondslag liggen van ons corpusanalytisch onderzoek:

- [HYP 3.1] Het Belgische Nederlands gebruikt minder vaak de rode volgorde [AUX+PART] dan het Nederlandse Nederlands (paragraaf 3.2.3.1);
- [HYP 3.2] In de dialecten komt de groene volgorde [PART+AUX] vaker voor dan de rode volgorde [AUX+PART] (paragraaf 3.2.3.2);
- [HYP 3.3] In vergelijking met de andere provincies in het Nederlandse taalgebied is de groene volgorde [PART+AUX] extreem populair in de noordelijke Nederlandse provincies (paragraaf 3.2.3.2);
- [HYP 3.4] In vergelijking met de andere provincies in het Nederlandse taalgebied is de rode volgorde [AUX+PART] erg populair in Nederlands-Limburg (paragraaf 3.2.3.2).

3.2.3 Resultaten en discussie

3.2.3.1 *Belgisch Nederlands versus Nederlands Nederlands*

Door een aantal historische en demografische ontwikkelingen, die elders uitvoerig beschreven zijn (Van der Wal & Van Bree 1994), heeft de standaardtaal in Vlaanderen en Nederland zich anders ontwikkeld: terwijl het Noorden een normaal standaardiseringstraject kon afleggen, is het Zuiden eeuwenlang van die standaardisering in het Noorden afgescheiden geweest en kon het door de druk van het Frans als elitetaal geen eigen standaard ontwikkelen. Sinds het einde van de negentiende eeuw heeft Vlaanderen opnieuw aansluiting gezocht met de standaard in Nederland, maar een gevolg van die verlate standaardisering is dat er in alle taalcomponenten verschillen zijn met het taalgebruik in Nederland (zie Geeraerts et al. 1999 voor een overzicht). Het Nederlands is daarmee lang niet de enige taal die verschillende regionale variëteiten heeft ontwikkeld (zie bv. de verschillen tussen Oostenrijks Duits en Duits Duits of de verschillen tussen Frans Frans en Canadees Frans), een fenomeen dat *pluricentriciteit* genoemd wordt. Vanuit die optiek is het dan ook niet vreemd om aan te nemen dat Vlaanderen en Nederland op een andere manier gebruik maken van de syntactische middelen die in de taal aanwezig zijn. Uit een

aantal macroperspectivische onderzoeken naar de regionale verspreiding van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde kwam inderdaad het volgende naar voren:

[HYP 3.1] Het Belgische Nederlands gebruikt minder vaak de rode volgorde [AUX+PART] dan het Nederlandse Nederlands.

Op het eerste gezicht lijken onze resultaten in tabel 3.4 die tendens te bevestigen:

Tabel 3.4: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per taalvariëteit.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Belgisch Nederlands	39.9% (3962/9920)	60.1% (5958/9920)
Nederlands Nederlands	32.6% (1747/5367)	67.4% (3620/5367)

Ondanks het feit dat de rode volgorde [AUX+PART] de vaakst gebruikte is in de beide delen van het taalgebied, tekent er zich toch een duidelijk verschil af tussen Vlaanderen en Nederland, in die zin dat de voorkeur voor de rode volgorde [AUX+PART] in het Belgische Nederlands minder sterk is dan in het Nederlandse Nederlands. Met behulp van een chi-kwadraattest wordt duidelijk dat het verschil tussen de beide regio's statistisch significant is ($\chi^2 = 81.26$, $df = 1$, $p < .0001$): dat betekent dat we voor 99.99% zeker zijn dat het verschil tussen de beide taalvariëteiten een reëel verschil is; de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde staat met andere woorden niet los van de regio waarin de volgorde geproduceerd wordt.

Problematisch aan het gebruik van de chi-kwadraattest is echter dat de grootte van de chi-kwadraatwaarde (en daarmee samenhangend de grootte van de p-waarde) beïnvloed wordt door de grootte van de dataset, in die zin dat, met behoud van dezelfde relatieve proporties, de chi-kwadraatwaarde op basis van een dataset met 1000 observaties 10 keer groter is dan op basis van 100 observaties. Concreet komt het

erop neer dat verschillen, hoe klein ook, in grote datasets heel snel statistisch significant worden. In dergelijke omstandigheden kan het daarom nuttig zijn om het effect in een betrouwbaarder perspectief te plaatsen door naar de waarden voor de odds ratio te kijken. De odds ratio (verder: o.r.) is een maat voor associatie die niet alleen stabiel blijft voor datasets van een verschillende grootteorde, maar die ook aangeeft of er een statistisch significant effect is en, indien dat effect vastgesteld kan worden, hoe sterk dat effect is op een schaal van 0 tot oneindig. Daarbij staat 1 voor geen associatie, minder dan 1 staat voor een negatieve associatie en meer dan 1 voor een positieve associatie. Hoe verder de o.r.-waarde verwijderd is van 1, hoe sterker de associatie. Zo zien we dat de waarschijnlijkheid van de groene volgorde [PART+AUX] (versus de rode volgorde [AUX+PART]) voor het Belgische Nederlands 38% groter is dan voor het Nederlandse Nederlands (o.r. = 1.38, B.I. = 1.29-1.48).⁵ Algemeen kunnen we dus besluiten dat de groene volgorde [PART+AUX] zonder meer frequenter gebruikt wordt in Vlaanderen dan in Nederland.

Nu kan men zich echter afvragen of het vastgestelde verschil tussen het Belgische en Nederlandse Nederlands geen schijneffect is dat teruggevoerd kan worden op een onderliggende, verborgen variabele. Bij de interpretatie van de resultaten van dergelijke bivariate statistische analyses moet daarom rekening gehouden worden met andere, niet-direct bestudeerde, covariërende verklarende variabelen die de relatie tussen de direct bestudeerde verklarende variabele en de responsvariabele (i.c. de woordvolgorde) beïnvloeden. Eén voorbeeld moet volstaan om het gevaar duidelijk te maken. Uit vorig onderzoek is gebleken dat de verdeling van de groene [PART+AUX] en rode [AUX+PART] volgorde verschilt per type hulpwerkwoord waarmee het deelwoord gecombineerd wordt: met name clusters met het hulpwerkwoord *zijn* hebben een grotere kans om in een groene volgorde [PART+AUX] voor te komen dan clusters met het hulpwerkwoord *hebben* (zie De Schutter 1964: 69; De Sutter et al. 2003: 118). Stel nu dat in het Belgisch-Nederlandse materiaal veel meer verbale eindgroepen met het hulpwerkwoord *zijn* geattesteerd zijn dan in het Nederlands-Nederlandse materiaal, dan zou het effect van regio gewoon een weerspiegeling kunnen zijn van de scheve verdeling tussen de verschillende types hulpwerkwoord over de twee genoemde regio's. Om het gevaar van overhaaste conclusies op basis van bivariate analyses te vermijden, is het daarom noodzakelijk dat

ook de conditionele associaties bestudeerd worden, i.e. de studie van het effect van een verklarende variabele op de respons, waarbij een derde variabele, de zgn. controlevariabele, constant gehouden wordt. In ons voorbeeld zou dat dan betekenen dat nagegaan wordt of er een effect is van regio op woordvolgorde als enkel gekeken wordt naar de eindgroepen met hulpwerkwoord *zijn* (bovenste luik van tabel 3.5); vervolgens wordt hetzelfde gedaan voor de hulpwerkwoorden *hebben* (middelste luik van tabel 3.5) en *worden* (onderste luik van tabel 3.5):

Tabel 3.5: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per taalvariëteit, gecontroleerd voor hulpwerkwoord.

		Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Zijn	Belgisch Nederlands	49.32% (1729/3506)	50.68% (1777/3506)
	Nederlands Nederlands	36.61% (748/2043)	63.39% (1295/2043)
Hebben	Belgisch Nederlands	31.72% (1013/3194)	68.28% (2181/3194)
	Nederlands Nederlands	27.65% (543/1964)	72.35% (1421/1964)
Worden	Belgisch Nederlands	37.89% (1220/3220)	62.11% (2000/3220)
	Nederlands Nederlands	33.53% (456/1360)	66.47% (904/1360)

Een van de meest gebruikte technieken om conditionele associaties te bestuderen, is de Cochran-Mantel-Haenszel-test (CMH), die onder meer een samenvatting geeft van de associatie tussen de verklarende variabele (i.c. regio) en de responsvariabele (i.c. woordvolgorde) over de verschillende partiële tabellen heen (i.e. de tabellen die gecontroleerd zijn voor een van de waarden van de controlevariabele, i.c. hulpwerkwoord). De CMH-analyse met controlevariabele *hulpwerkwoord* laat zien dat er een globaal significant effect is van regio op woordvolgorde over alle strata heen (*zijn* vs. *hebben* vs. *worden*; $CMH-\chi^2 = 81.64$, $df = 1$, $p < .0001$). Dat resultaat geeft aan dat het regionale effect dat we in de marginale tabel 3.4 zagen, stabiel blijft voor

de verschillende hulpwerkwoorden. Zoals gezegd vatten de globale CMH-statistieken de informatie uit de verschillende partiële tabellen samen. Daarbij kan het voorvallen dat een globaal significant resultaat het gevolg is van een sterk significant effect in een van de partiële tabellen, terwijl de andere partiële tabellen geen of een minder significant effect vertonen; in dat geval spreken we van een interactie. Vanuit die optiek is het raadzaam om telkens ook een gedetailleerdere analyse van de verschillende strata uit te voeren om interactie-effecten op het spoor te komen. In dit geval bleken er geen interactie-effecten te zijn (in het vervolg van dit proefschrift zullen we ervan afzien om bij elke CMH-analyse ook de conditionele tabel, zoals tabel 3.5, te geven; met behulp van de datamatrix die we via <http://www.ling.arts.kuleuven.ac.be/qlvl/datasets> ter beschikking stellen, kunnen de tabellen echter gemakkelijk zelf gegenereerd worden). Voor de CGN-data (R5) is vervolgens ook nog nagegaan of het regionale effect niet teruggevoerd kan worden op een van de volgende contextuele controlevariabelen: geslacht (man vs. vrouw), leeftijd (> 44 vs. < 45). Dat bleek niet het geval te zijn: geslacht: $\text{CMH-}\chi^2 = 16.20$, $\text{df} = 1$, $p < .0001$; leeftijd: $\text{CMH-}\chi^2 = 29.64$, $\text{df} = 1$, $p < .0001$; er waren geen interactie-effecten.

Samenvattend kunnen we zeggen dat er een reëel verschil lijkt te zijn tussen het Belgische en het Nederlandse Nederlands in het gebruik van de verschillende woordvolgordevarianten. Daarmee bevestigen we [HYP 3.1] en, meer bepaald, de bevindingen van Verhasselt (1961), Stroobants (1997) en De Schutter (1996: 213); de resultaten van Haeseryn (1990) kunnen niet worden bevestigd. Diachroon onderzoek zal moeten uitmaken of het verschil tussen Vlaanderen en Nederland een relict is van de specifieke historische ontwikkeling van de standaardtaal in de beide gebieden, waarbij de hypothese is dat de vastgestelde lexicaal-onomasiologische toenadering van Vlaanderen naar Nederland (Geeraerts et al. 1999) al dan niet in vertraagde vorm ook van toepassing is op de solidere, abstractere taalcomponent syntaxis, dan wel of het regionale verschil teruggevoerd kan worden op een pragmatisch-functioneel verschillend gebruik van de syntactische variatie in de beide gebieden.⁶

3.2.3.2 *Streekverschillen*

In deze paragraaf proberen we na te gaan of in onze dataset sporen te vinden zijn van wat uit dialectologisch onderzoek gebleken is, met name:

- [HYP 3.2] De algemene voorkeur voor de groene volgorde [PART+AUX] in de dialecten;
- [HYP 3.3] De extreme voorkeur van de noordelijke provincies Friesland, Groningen en Drenthe voor de groene volgorde [PART+AUX];
- [HYP 3.4] De meer dan gemiddelde voorkeur van Nederlands Limburg voor de rode volgorde [AUX+PART].

Omdat de observaties in onze dataset enkel voor het gesproken materiaal in het CGN (R5) gediversifieerd zijn naar de verschillende streken en provincies, laten we het geschreven materiaal van CONDIV buiten beschouwing. Daarnaast moet er rekening mee gehouden worden dat we niet beschikken over een corpus van dialectologisch materiaal, maar enkel een aantal gesproken teksten die al dan niet spontaan tot stand gekomen zijn. Vertrekkend van de algemene hypothese dat taalgebruikers in de minst formele communicatiesituaties het nauwst aanleunen bij de syntaxis van het dialect waarmee ze opgegroeid zijn, zullen we daarom enkel de privédialogen, die spontaan en in een niet-openbare context tot stand gekomen zijn, in de analyse betrekken. Tabel 3.6 geeft de verdeling van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde over de verschillende provincies in het Nederlandse taalgebied (n = 1128; Belgisch Nederlands: n = 418; Nederlands Nederlands: n = 710):

Tabel 3.6: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per provincie.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Vlaams-Brabant, Antwerpen	68.8% (97/141)	31.2% (44/141)
Oost-Vlaanderen	93.3% (70/75)	6.7% (5/75)
West-Vlaanderen	80.7% (71/88)	19.3% (17/88)
Belgisch Limburg	84.2% (96/114)	15.8% (18/114)

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Groningen, Drenthe, Friesland	42.2% (19/45)	57.8% (26/45)
Overijssel	66.0% (33/50)	34.0% (17/50)
Gelderland	54.8% (85/155)	45.2% (70/155)
Utrecht	49.0% (24/49)	51.0% (25/49)
Noord-Holland	61.0% (25/41)	39.0% (16/41)
Zuid-Holland	57.3% (47/82)	42.7% (35/82)
Zeeland	26.9% (7/26)	73.2% (19/26)
Noord-Brabant	58.0% (87/150)	42.0% (63/150)
Nederlands Limburg	51.8% (58/112)	48.2% (54/112)

In wat volgt zetten we de belangrijkste vaststellingen op een rijtje.

Algemeen

In tegenstelling tot het overzicht van de rode volgorde [AUX+PART] in de hele dataset, vertonen de gesproken privédialogen globaal genomen een (grotere) voorkeur voor de groene volgorde [PART+AUX] (over alle provincies heen: 64%; over alle Vlaamse provincies heen: 80%; over alle Nederlandse provincies heen: 54%; zie tabel 3.6). Dat zou erop kunnen wijzen dat de dominantie van de rode volgorde [AUX+PART] die in de inleiding tot dit hoofdstuk vastgesteld werd vooral te wijten is aan de dominantie van de formele en/of de geschreven communicatie in onze dataset (zie paragraaf 3.3). De discrepantie is in ieder geval een gedeeltelijke bevestiging van wat uit dialectologisch onderzoek al duidelijk geworden was, met name het primaat

van de groene volgorde [PART+AUX] in gesproken, niet-publieke communicatieomstandigheden ([HYP 3.2]).

Vlaanderen

De voor de hele dataset vastgestelde grotere voorkeur in Vlaanderen voor de groene woordvolgorde [PART+AUX] lijkt op basis van het materiaal in het subcorpus van gesproken privédialogen bevestigd te worden, en wel voor alle Vlaamse provincies. De perifere gebieden Oost- en West-Vlaanderen en Belgisch Limburg vertonen op zijn minst 80% groene woordvolgordes [PART+AUX], enkel de Brabantse regio sluit met 68.8% groene volgordes [PART+AUX] aan bij de meest groene regio in Nederland, Overijssel (66%). De verschillen tussen de onderscheiden regio's in Vlaanderen zijn statistisch hoogsignificant ($\chi^2 = 20.61$, $df = 3$, $p = .0001$).

Om de aard van de associatie tussen regio en woordvolgorde beter te begrijpen, voeren we een residu-analyse en een gepartitioneerde chi-kwadraatanalyse uit. Met behulp van een residu-analyse gaan we na of er een bepaalde regio aangewezen kan worden die meer dan de andere regio's heeft bijgedragen tot de globale sterkte van het effect. In ons geval blijken dat voornamelijk Oost-Vlaanderen en de Brabantse regio te zijn: in Oost-Vlaanderen (residu = 3.2) wordt de groene volgorde [PART+AUX] en in de Brabantse regio (residu = 4.04) de rode volgorde [AUX+PART] vaker dan verwacht gebruikt. West-Vlaanderen en Limburg wijken niet af van wat algemeen gebruikelijk is. Uit de gepartitioneerde chi-kwadraatanalyse, die erop gericht is om het aantal categorieën in een tabel met meer dan 2 categorieën te reduceren door groepen samen te nemen die statistisch niet-significant van elkaar verschillen, blijkt hetzelfde verschil tussen West-Vlaanderen en Limburg versus de Brabantse regio versus Oost-Vlaanderen.

Om na te gaan of het regionale effect in het Belgische Nederlands niet te wijten is aan een onderliggende variabele, hebben we het effect van de regionale variabele geanalyseerd terwijl de variabelen *geslacht*, *type gesprek* (conversatie of interview; andere types gesprekken waren niet voorhanden), *leeftijd* (jong: 18-34 jaar; oud: >34 jaar) en *hulpwerkwoord* (zie De Schutter 1964: 69; De Sutter et al. 2003: 118) constant gehouden werden. Controleren voor opleidingsniveau bleek overbodig, aangezien de meeste werkwoordelijke eindgroepen afkomstig zijn van hoogopgeleide sprekers (407

op 418 attestaties; de 11 resterende attestaties werden geproduceerd door sprekers van het middelste opleidingsniveau; er zijn geen eindgroepen voor het laagste opleidingsniveau); hetzelfde geldt voor het beroepsniveau (354 op 418 attestaties zijn gelokaliseerd in de groep met opleidingsniveau C, onderwijzend en academisch personeel).⁷ De globale resultaten van de conditionele CMH-analyses bevestigen de eerder verkregen resultaten, i.e. er is een significant regionaal effect op de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde, ook als gecontroleerd wordt voor *geslacht* (CMH- $\chi^2 = 18.12$, $df = 3$, $p = .0004$), *type gesprek* (CMH- $\chi^2 = 21.28$, $df = 3$, $p < .0001$), *leeftijdscategorie* (CMH- $\chi^2 = 12.82$, $df = 3$, $p = .0051$) en *hulpwerkwoord* (CMH- $\chi^2 = 12.69$, $df = 3$, $p = .0001$). Het regionale effect bleef behouden over alle strata van alle controlevariabelen heen, behalve voor *hulpwerkwoord*, waar een interactie-effect met *hebben* kon worden vastgesteld (*zijn*: $\chi^2 = 3.68$, $df = 2$, $p > .05$; *hebben*: $\chi^2 = 22.86$, $df = 2$, $p < .0001$; *worden*: meer dan 25% van de cellen bevatten minder dan 5 attestaties, daarom kan geen betrouwbare chi-kwadraatanalyse uitgevoerd worden).⁸ Het interactie-effect tussen de variabelen *regio* en het hulpwerkwoord *hebben* geeft aan dat de verschillende Vlaamse provincies significant van elkaar verschillen in de keuze voor een van de volgordevarianten in werkwoordelijke eindgroepen met het hulpwerkwoord *hebben*. De clusters met *hebben* zijn, met andere woorden, verantwoordelijk voor de verschillen tussen de verschillende provincies, terwijl het gebruik van de syntactische varianten in eindgroepen met *zijn* (voor *worden* kunnen we dat niet nagaan, wegens onvoldoende grote celfrequenties) homogeen is voor het hele Belgisch-Nederlandse gebied.

Samenvattend kunnen we zeggen dat de Vlaamse provincies, ondanks een algemene voorkeur voor de groene volgorde [PART+AUX], op het vlak van de hier bestudeerde syntactische preferenties geen homogeen gebied zijn: het Brabantse materiaal bevat vaker dan gemiddeld rode volgordes [AUX+PART] en het Oost-Vlaamse materiaal vaker groene volgordes [PART+AUX]. Vooral dat laatste resultaat is interessant tegen de achtergrond van Haeseryns (1990) vaststelling dat West-Vlamingen een voorkeur lijken te hebben voor de groene volgorde [PART+AUX] in vergelijking met de centrale Brabants-Antwerpse regio. Zijn vaststelling wordt door ons onderzoek bevestigd, maar tegelijk ook genuanceerd, aangezien de Oost-Vlamingen een nog sterkere voorkeur voor de groene volgorde [PART+AUX] lijken te hebben. Dat de Brabantse regio vaker de rode volgorde [AUX+PART] gebruikt, en

daarmee aansluit bij wat in Nederland gebruikelijk is, kan met de politieke, sociaal-economische en daarmee linguïstische centrumfunctie te maken hebben; als het inderdaad zo is, zoals in de sociolinguïstische literatuur gezegd wordt, dat de rode volgorde [AUX+PART] de prestigevariant is (Sassen 1963, Sturm 1990), dan kan verwacht worden dat die variant frequenter gebruikt wordt in Brabant als leidinggevende regio. Aannemen dat hier sprake zou zijn van een Nederlands-Nederlandse invloed op het Belgische Nederlands—het Belgisch Nederlands groeit naar het Nederlandse Nederlands toe (Geeraerts et al. 1999) en daarbij neemt de centrale, Brabantse regio de leiding—is voorbarig, maar moet zeker tot de mogelijkheden gerekend worden. Verder onderzoek zal hierin klaarheid moeten scheppen.

Nederland

Complementair met wat hierboven voor Vlaanderen gezegd is, is de vaststelling dat in Nederland over het algemeen vaker de rode volgorde [AUX+PART] wordt gebruikt. Verder vallen vooral de extreme voorkeuren van Zeeland (73.1% rode volgordes [AUX+PART]) en Overijssel (66% groene volgordes [PART+AUX]) op. De verschillen tussen de onderscheiden regio's zijn weliswaar significant ($\chi^2 = 15.98$, $df = 8$, $p = .04$), maar wel een heel stuk minder dan de verschillen tussen de Vlaamse regio's, hetgeen erop kan wijzen dat het Nederlandse Nederlands homogener is dan het Belgische Nederlands.

Uit de residu-analyse en de gepartitioneerde chi-kwadraattest blijkt dat het de extreme positie van het Zeeuwse materiaal is die voor het significante effect zorgt. Gegeven het kleine aantal attestaties voor Zeeland ($n = 26$) kan men terecht vraagtekens plaatsen bij dat resultaat.

Verder blijkt uit tabel 3.6 dat niets van de eerder vastgestelde tendensen in onze dataset teruggevonden kan worden: de noordelijke provincies vertonen in geen geval een dominante voorkeur voor de groene volgorde [PART+AUX] (falsificatie van [HYP 3.3]) en Nederlands Limburg heeft geen significant extremer rood-gebruik [AUX+PART] dan andere regio's in Nederland (falsificatie van [HYP 3.4]). De reden voor die discrepantie kan er in de eerste plaats mee te maken hebben dat we niet over een corpus van dialectmateriaal beschikken en dus op geen enkele wijze direct kunnen

vergelijken met wat eerder op dialectologisch niveau vastgesteld is. In de tweede plaats moet ook opgemerkt worden dat de absolute frequenties per provincie niet erg hoog zijn, wat maakt dat we de verkregen resultaten met enige omzichtigheid moeten behandelen.⁹

3.3 De impact van register

In deze sectie willen we nagaan wat de invloed van de communicatieve situatie—of kortweg register—is op de gekozen woordvolgorde in de tweeledige werkwoordelijke eindgroep. Eerst geven we een overzicht van de belangrijkste literatuur ter zake (paragraaf 3.3.1), vervolgens worden de verschillende hypotheses voorgesteld (paragraaf 3.3.2) en tot slot presenteren en evalueren we de resultaten van ons onderzoek (paragraaf 3.3.3), dat het effect van register langs drie intergerelateerde dimensies bestudeert: modus (paragraaf 3.3.3.1), mate van interactie (paragraaf 3.3.3.2) en mate van redactionele controle (paragraaf 3.3.3.3). In een slotparagraaf (3.3.3.4) vergelijken we de verschillende dimensies onderling.

3.3.1 *Status quaestionis*

Er bestaat een grote hoeveelheid aan wetenschappelijk onderzoek waaruit blijkt dat taalgebruikers in het taalproductieproces de middelen die de taal hen ter beschikking stelt op een intrigerend systematische manier (al dan niet bewust) uitkiezen in functie van de communicatieve situatie waarin ze zich bevinden. Een van de meest opvallende omstandigheden van een dergelijk keuzeprocess is het ogenblik waarop een taalgebruiker binnen een zelfde communicatieve omgeving plotseling van de ene variëteit van een taal op een andere variëteit van dezelfde taal overschakelt als een antwoord op bepaalde communicatief-situationele stimuli en/of als een strategische keuze om sociale betekenis te genereren (zgn. *style shifting*; Milroy & Gordon 2003: 198-222). Het hoeft dan ook niet te verbazen dat het precies in het domein van de taalvariatie is, het domein waar taalgebruikers over twee of meer concurrerende vormen beschikken voor dezelfde conceptuele inhoud, dat taalwetenschappers veel onderzoek gedaan hebben naar de invloed van de communicatieve situatie.

Ook de impact van verschillende communicatieve situaties op de woordvolgorde in de eindgroep is niet onopgemerkt voorbijgegaan in de literatuur. Volgens de ANS (1997: 1067-1069) “[...] tekent er zich een verschil tussen gesproken en geschreven taal af. De gebruikelijkste volgorde in gesproken taal, zeker in minder formeel taalgebruik, is die met het deelwoord eerst [de groene volgorde [PART+AUX], GDS]. In geschreven taal bestaat er een tendens om vaker de omgekeerde volgorde te gebruiken [de rode volgorde [AUX+PART], GDS]. Die tendens is het duidelijkst in journalistieke teksten”. Het onderscheid langs twee intergerelateerde dimensies, gesproken versus geschreven en formeel versus informeel, vinden we terug in de wetenschappelijke literatuur. Zo stelt Pauwels (1953) vast dat de rode volgorde [AUX+PART] vaker gebruikt wordt in de schrijftaal dan in de gesproken volkstaal en merken Stroop (1970) en De Rooij (1986) dat in krantentaal opvallend vaker de rode volgorde [AUX+PART] voorkomt dan in (literaire en) gesproken taal. Sassen (1963), De Hoop & Smabers (1987) en De Schutter (1996) komen tot dezelfde vaststelling (zie ook de taaladviesliteratuur in 1.3.1.2). Sassen voegt er nog aan toe dat diezelfde neiging tot rood [AUX+PART] ook in ambtelijk en wetenschappelijk proza te vinden is, De Schutter (1996) verwijst naar de voorkeur voor de rode volgorde [AUX+PART] van vertalers. Die laatste waarschuwt er in datzelfde artikel ook nog voor dat de twee dimensies niet op een hoopje mogen worden gegooid: geschreven en formeel lijken op het eerste gezicht in hoge mate met elkaar samen te vallen, net als gesproken en informeel lijken te correleren, maar in werkelijkheid hoeft dat niet altijd het geval te zijn. In vorig onderzoek heeft men daarom soms te snel geconcludeerd dat de groene woordvolgorde [PART+AUX] vaker voorkomt in de gesproken taal en de rode volgorde [AUX+PART] vaker in de geschreven taal, terwijl er in feite nog een andere, onderliggende, verklarende variabele in de analyse betrokken had moeten worden, namelijk formaliteit van het register (zie De Rooij 1986, Stroobants 1997, Zuckerman 2001). De Schutter (1996) is zich van die beide dimensies bewust en classificeert televisiegesprekken en televisie-interviews dan ook als formele gesproken taal tegenover informele gesproken taal en formele en informele geschreven taal (zie ook Sassen 1963 die opmerkt dat de voorkeur voor de rode volgorde [AUX+PART] ook in gesproken radioberichten terug te vinden is). Nu kan men terecht opmerken dat de identificatie van de beide dimensies in het registeronderzoek op zich ook een idealisatie is van de werkelijkheid en dat er in feite nog meer dimensies onderscheiden moeten worden. Het is inderdaad zo dat een

concept als ‘formaliteitsgraad’ op zich in verschillende componenten kan worden ontleed. We zullen bij de bespreking van de resultaten hierop terugkomen.

Niet zelden wordt het aangehaalde registersverschil verklaard door een anti-dialectreflex: zoals uit het bovenstaande overzicht duidelijk geworden is, is de groene woordvolgorde [PART+AUX] de meest gebruikelijke in de dialecten. Veel wetenschappers gaan ervan uit dat taalgebruikers het gebruik van dialect associëren met het volkse, onverzorgde, onbeschaafde karakter van het dorpsleven en zich in formelere situaties al dan niet bewust willen afzetten tegen wat in het dialect gangbaar is (Sassen 1963, De Schutter 1996, Van der Horst & Van der Horst 1999). Sassen (1963) en Sturm (1990) spreken in die context over de rode volgorde [AUX+PART] als prestigeform en de groene volgorde [PART+AUX] als gestigmatiseerde vorm (zie ook de literatuurbespreking in paragraaf 1.3.1.2), maar uit de literatuurbespreking van het onderzoek dat zich met de oordelen van taalgebruikers beziggehouden heeft, blijkt dat dat onderscheid zeker niet voor alle geledingen van de taalgemeenschap geldt (paragraaf 1.3.5). Overigens is niet iedereen het eens over de invloed die het register zou uitoefenen op de keuze voor een van de woordvolgordes: Zuckerman (2001: 34) meent dat het register vroeger een rol gespeeld heeft, maar niet meer in het huidige Nederlands. Hij ondersteunt die claim door te verwijzen naar het onderzoek van De Hoop & Smabers (1987) (waarbij opgemerkt moet worden dat dat onderzoek peilde naar taaloordeelen en niet naar reëel taalgedrag) en een eigen vertaalexperiment (waarvan het design niet toegelicht wordt).

3.3.2 *Hypotheses*

In wat volgt zullen we de bevindingen van Zuckerman en alle andere wetenschappers die onderzoek gedaan hebben naar de invloed van het register op de woordvolgorde confronteren met de patronen in onze data. De twee hypothesen die aan de basis zullen liggen van ons eigen onderzoek zijn de volgende:

- [HYP 3.5] In geschreven taal komt de rode volgorde [AUX+PART] vaker voor dan in gesproken taal (paragraaf 3.3.3.1);
- [HYP 3.6] In formeel taalgebruik komt de rode volgorde [AUX+PART] vaker voor dan in informeel taalgebruik (paragraaf 3.3.3.2 en 3.3.3.3).

3.3.3 Resultaten en discussie

Vooraleer we beginnen met de bespreking van de resultaten van ons onderzoek, staan we eerst nog even stil bij de manier waarop de factor *register* geoperationaliseerd zal worden. Uit het bovenstaande literatuuroverzicht is gebleken dat het register in vroegere studies vaak op een beperkte, dichotome manier geoperationaliseerd is, met name als het onderscheid tussen geschreven en gesproken taal of tussen krantentaal en literaire taal. Wanneer men in aanmerking neemt dat in heel wat van die onderzoeken betrekkelijk wat onduidelijkheid bestaat over het type materiaal dat gebruikt is en de dataselectiecriteria die gehanteerd werden, en dat het daarom erg moeilijk is om de resultaten die van verschillende studies direct met elkaar te vergelijken (zie de kritiek die we geformuleerd hebben in paragraaf 1.3.4), is het makkelijk in te zien dat het nog steeds niet geheel duidelijk is wat bv. de (relatieve) positie is van geschreven formeel, geschreven informeel, gesproken formeel en gesproken informeel taalgebruik op het vlak van de gekozen woordvolgorde.

Op basis van onze dataset, die bestaat uit geschreven en gesproken taal, uit formeel en informeel taalgebruik zullen we trachten die lacune aan te vullen. Om een beter (i.e. gediversifieerder en genuanceerder) zicht te krijgen op de distributie van de rode [AUX+PART] en de groene [PART+AUX] woordvolgorde in verschillende types register en om na te gaan of de keuze voor woordvolgorde afhankelijk is—en zo ja hoe sterk—van het register, zullen we meer bepaald de acht subcomponenten van ons corpus (zie hoofdstuk 2) gebruiken: publieke monologen, publieke dialogen, privédialogen, IRC-materiaal, Usenet-materiaal, regionale populaire kranten, nationale populaire kranten en kwaliteitskranten. Die componenten kunnen volgens een aantal dimensies ingedeeld worden (zie bv. Biber 1988: 199-200):

- *Modus*: geschreven (CONDIV-materiaal) versus gesproken (CGN-materiaal (R5)) taal.
- *Interactiviteit*: eenrichtingscommunicatie (monologen en krantenmateriaal) versus twee- of meer-richtingscommunicatie (dialogen, Usenet- en IRC-materiaal).
- *Redactionele controle*: weinig redactionele controle (IRC en dialogen) versus beperkte redactionele controle (monologen en Usenet) versus geavanceerde redactionele controle (krantenmateriaal).

Het is duidelijk dat de genoemde dimensies met elkaar interageren, zoals bijvoorbeeld de dimensies *interactiviteit* en *redactionele controle*: over het algemeen is het zo dat hoe groter de interactiviteit is, hoe geringer de redactionele controle. Het doel van deze indeling is dan ook niet een discrete behandeling van de verschillende dimensies, maar een vergelijking van de samenhang en verschillen tussen die dimensies.

In de rest van deze sectie worden de drie dimensies apart bestudeerd en vervolgens in een globale analyse met elkaar vergeleken.

3.3.3.1 Gesproken versus geschreven taal (modus)

Tabel 3.7 vat de gegevens samen van het gebruik van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde in functie van de eerste dimensie, i.e. het gesproken versus het geschreven Nederlands. We brengen daarbij eerst nog [HYP 3.5] in herinnering:

[HYP 3.5] In geschreven taal komt de rode volgorde [AUX+PART] vaker voor dan in gesproken taal.

Tabel 3.7: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per modus.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Gesproken Nederlands	49.1% (1913/3894)	50.9% (1981/3894)
Geschreven Nederlands	33.3% (3796/11393)	66.7% (7579/11393)

In eerste instantie is het relatief grote aantal verbale clusters in het gesproken materiaal opvallend. Ondanks het feit dat de hoeveelheid gesproken materiaal waarover we beschikken meer dan 10 keer kleiner is dan de hoeveelheid geschreven materiaal, en de kans op een verbale cluster dus zowat 10 keer kleiner is, zijn er slechts 2.9 keer minder verbale clusters gevonden in het gesproken materiaal in vergelijking met het geschreven materiaal. In tegenstelling tot wat men intuïtief zou verwachten, zou dat

erop kunnen wijzen dat het voorkomen van een werkwoordelijke eindgroep in een complementzin een marker is voor gesproken taal en niet voor de schijnbaar gesofisticeerdere geschreven taal. Die hypothese wordt bevestigd door het onderzoek dat Biber (1988: 113-114) heeft uitgevoerd naar verschillende dimensies in gesproken en geschreven taal:

The distributional pattern shown [...] runs counter to previous theoretical expectations: several subordination measures that are typically associated with informational elaboration co-occur here with demonstratives, final prepositions, and demonstrative pronouns, which are associated with informal unplanned types of discourse. [...] suggests that they function to mark informational elaboration in relatively unplanned types of discourse [...]. [...] that is, because spoken language is produced and comprehended as an on-going process, it is characterized by 'an intricacy of movement [and by] complex sentence structures with low lexical density [...]'. [They] apparently mark informational elaboration that is produced under strict real-time constraints, resulting in a fragmented presentation of information accomplished by tacking on additional dependent clauses, rather than an integrated presentation that packs information into fewer constructions containing more high-content words and phrases [...]. In addition that complements to verbs and adjectives can be used for elaboration of information relative to the personal stance of the speaker, introducing an affective component into this dimension (e.g., I wish that...[...]).

Dat complementzinnen met *dat* vaker gebruikt worden in gesproken taal hoeft op zich uiteraard nog niet te betekenen dat er ook meer tweeledige werkwoordelijke eindgroepen in gesproken taal geproduceerd worden, maar het spreekt voor zich dat een hogere frequentie van complementzinnen een verhoogde kans op tweeledige werkwoordelijke eindgroepen met zich brengt.¹⁰

Tabel 3.7 laat verder nog zien dat in de gesproken taal, zoals verwacht ([HYP 3.5]), vaker de groene woordvolgorde [PART+AUX] gebruikt wordt dan in de geschreven taal. Het verschil tussen de beide modi is statistisch significant ($\chi^2 = 307.61$, $df = 1$, $p < .0001$). De significante o.r. van 1.93 (B.I. = 1.79-2.08) wijst erop dat de waarschijnlijkheid van een groene woordvolgorde [PART+AUX] tegenover de

waarschijnlijkheid van een rode volgorde [AUX+PART] in de gesproken taal bijna twee keer hoger ligt dan in de geschreven taal, of anders, dat de geschatte odds van de groene woordvolgorde [PART+AUX] voor het gesproken Nederlands gelijk is aan ca. 2 keer de geschatte odds van de groene volgorde [PART+AUX] in de geschreven taal.¹¹

Net zoals bij de bestudering van het regionale effect verdient het ook hier aanbeveling om de conditionele associaties te bestuderen. Uit de CMH-analyse blijkt dat het effect van modus stabiel blijft indien men controleert voor hulpwerkwoord (*zijn* vs. *hebben* vs. *worden*; CMH- $\chi^2 = 331.53$, $df = 1$, $p < .0001$)¹²; er zijn geen redundantie- of interactie-effecten.

De vraag die we ons nu nog kunnen stellen, is: waarom hangt de woordvolgordevoorkeur af van de modus van het taalgebruik? Een mogelijke verklaring voor dat effect zou gevonden kunnen worden in de verschillende productie-eisen die gepaard gaan met de gesproken en geschreven taalproductie, i.e. in gesproken taal wordt de taalgebruiker blootgesteld aan tijdsbeperkingen op de fonologische en prosodische encoding van de geïntendeerde boodschap¹³, terwijl die tijdsbeperkingen in geschreven taal volledig afwezig zijn. Dat verschil in productie-eisen zou kunnen verklaren waarom in de gesproken taal vaker dan in de geschreven taal teruggerepen wordt naar de groene woordvolgorde [PART+AUX], die, blijkens de hierboven gepresenteerde dialectologische studies de meest basale is, i.e. het frequentst in de meest basale taalvariëteit voor de meest basale communicatiesituaties gebruikt wordt. Voorlopig is dat uiteraard enkel een hypothese. In het verdere verloop van dit hoofdstuk en dit proefschrift komen we nog uitgebreid op die productiehypothese terug.

3.3.3.2 *Mate van interactie*

Langs een tweede dimensie werd het materiaal dat voor deze studie gebruikt wordt, onderscheiden naar mate van interactiviteit. Ook al lijkt deze dimensie volledig te overlappen met de vorige dimensie—geschreven correleert met eenrichtingscommunicatie, gesproken met tweerichtingscommunicatie—toch snijdt de dimensie interactiviteit dwars door de vorige dimensie heen: zowel gesproken publieke monologen als geschreven krantenartikelen zijn vormen van

eenrichtingscommunicatie; IRC-logs hebben het dialogische karakter gemeen met de CGN-dialogen (R5) en komen daarom in dezelfde groep van tweerichtingscommunicatie terecht. Als we even aannemen dat de mate van interactie invers samenhangt met de graad van formaliteit (hoe hoger de interactie, hoe lager de formaliteit), dan kunnen we in deze paragraaf [HYP 3.6] testen:

[HYP 3.6] In formeel taalgebruik komt de rode volgorde [AUX+PART] vaker voor dan in informeel taalgebruik.

Tabel 3.8 geeft een overzicht van het gebruik van de twee woordvolgordes, onderscheiden naar type interactie:

Tabel 3.8: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type interactie.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
1-richtingscommunicatie	28.8% (3156/10966)	71.2% (7810/10966)
2-richtingscommunicatie	59.1% (2553/4321)	40.9% (1768/4321)

Uit de tabel blijkt duidelijk dat de aanwezigheid van interactie tussen twee of meer gesprekspartners het gebruik van de groene woordvolgorde [PART+AUX] bevordert ($\chi^2 = 1216.50$, $df = 1$, $p < .0001$). De significante o.r. van 0.28 (B.I. = 0.26-0.30) maakt duidelijk dat de waarschijnlijkheid van de groene woordvolgorde [PART+AUX] versus de waarschijnlijkheid van de rode woordvolgorde [AUX+PART] zowat vier keer kleiner is in interactieloze communicatiesituaties dan in interactieve omstandigheden (bevestiging van [HYP 3.6]). Ook hier geven de resultaten van de conditionele CMH-analyses aan dat het effect stabiel blijft indien men controleert voor *hulpwerkwoord* (*zijn vs. hebben vs. worden*; $CMH-\chi^2 = 1293.39$, $df = 1$, $p < .0001$), *geslacht* (*m vs. v*; $CMH-\chi^2 = 238.26$, $df = 1$, $p < .0001$) of *leeftijdscategorie* (*> 44 vs. < 45*; $CMH-\chi^2 = 224.81$, $df = 1$, $p < .0001$)¹⁴; er zijn geen redundantie- of interactie-effecten.

Ondanks het feit dat deze dimensie niet volledig gelijkloopt met de vorige dimensie, vertonen de beide dimensies wel een sterke verwantschap op het vlak van de productieomstandigheden: het verschil tussen geschreven en gesproken taal hebben we geïnterpreteerd als een verschil in productiedruk op articulatorisch vlak. Traditioneel bestaat de neiging om het verschil tussen geschreven en gesproken taal ook gelijk te stellen met een algemeen verschil in productie-eisen, i.e. in de geschreven modus kan de taalgebruiker rustig de tijd nemen voor de grammaticale encoding van de geïntendeerde boodschap (lexicale keuze, functietoekenning en woordvolgorde; Bock & Levelt 1994: 946) en heeft de taalgebruiker de kans om achteraf de geschreven tekst nog een keer te herzien, in gesproken taal moet de taalgebruiker diezelfde keuzes maken op een veel beperktere tijd en heeft hij niet meer de mogelijkheid om later terug te komen op eerder gemaakte keuzes (in die zin dat de uiting volledig gesubstitueerd kan worden door de aangepaste versie). Echter, die productiedruk in gesproken taal kan in sommige omstandigheden verminderd worden door voorbereidend werk, zoals het geval is in toespraken en nieuwsberichten. Het spreekt voor zich dat de productiedruk op articulatorisch vlak onverminderd blijft, maar de lexicale en morfosyntactische keuzes zijn over het algemeen op voorhand gemaakt en zijn in die zin vergelijkbaar met de taalproductieomstandigheden die gelden voor de geschreven modus. Omgekeerd kan de productiedruk in geschreven taal stijgen door een communicatiesituatie die volledig lijkt op een gesprek tussen twee of meer gesprekspartners, met als enige verschil de modus waarin de communicatie verloopt. Een voorbeeld van een dergelijke communicatie vinden we in IRC-kanalen. IRC (Internet Relay Chat) is een internetmodule waarmee men on line en simultaan met elkaar kan converseren; door het simultane en hoog-interactieve karakter van de communicatievorm stijgt de productiedruk zoals in spontane gesproken dialogen (wat zich bv. uit in korte, elliptische zinnen; Ko 1996).

In de tweede dimensie worden de communicatievormen met een hevige productiedruk gegroepeerd en afgezet tegen de communicatievormen met geen of een verminderde productiedruk: in de eenrichtingscommunicatie wordt de productiedruk tot een minimum herleid doordat het eindproduct het resultaat is van voorbereidend werk (de mate van inmenging van de communicatiepartners tijdens de communicatie is minimaal tot onbestaand), in de tweerichtingscommunicatie blijft de productiedruk overeind, omdat de taalgebruiker moet kunnen inspelen op wat de

communicatiepartner zegt of vraagt. Wegens de grote mate van onverwachtheid, kan een eventuele voorbereiding van het eigen discours in die omstandigheden maar beperkt verlichtend werken.

Bij een vergelijking van de twee dimensies (tabel 3.7 en 3.8) valt het op dat de tweede dimensie, met het verschil tussen aan- en afwezigheid van interactie en daarmee aan- of afwezigheid van globale productiedruk, een grotere discrepantie oplevert dan de eerste dimensie, waar gefocust wordt op het verschil tussen gesproken en geschreven taal en dus op het verschil tussen al dan niet aanwezigheid van articulatorische productiedruk. Die grotere discrepantie uit zich onder meer in een grotere chi-kwadraatwaarde (307.61 voor de eerste dimensie versus 1216.5 voor de tweede dimensie) en o.r.-waarde (1.93 voor de modus en $(1 / 0.28 =) 3.57$ voor mate van interactie). Dat het effect van de tweede dimensie groter is dan het effect van de eerste dimensie, hoeft uiteraard niet te betekenen dat de eerste dimensie als irrelevant beschouwd moet worden, maar het lijkt wel alsof de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde niet zozeer door de articulatorische productiedruk bepaald wordt, maar door de globale productiedruk waaraan taalgebruikers blootgesteld worden: een hoge productiedruk levert meer groene woordvolgordes [PART+AUX] op, een lage productiedruk levert meer rode woordvolgordes [AUX+PART] op. Bij de vergelijking van de drie dimensies verderop komen we op de precieze verhouding van de eerste en tweede dimensie terug.

3.3.3.3 *Mate van redactionele controle*¹⁵

Zoals gezegd hangt de mate van interactiviteit nauw samen met de mate van productiedruk: hoe hoger de interactie, hoe hoger de productiedruk. Daaraan kunnen we nu de mate van redactionele controle toevoegen: hoe hoger de interactie, hoe hoger de productiedruk en hoe minder controle je kan uitoefenen op het resultaat van het productieproces. Op het eerste gezicht lijkt het er dus op dat de mate van redactionele controle volledig overlapt met de mate van interactiviteit. Er is echter één verschil tussen de beide dimensies, met name een verdere opsplitsing van het aantal categorieën. In de tweede dimensie was er sprake van aan- of afwezigheid van interactie in de communicatie, in de derde dimensie maken we het onderscheid tussen weinig redactionele controle (correleert met veel interactie: IRC-dialogen en

gesproken dialogen), beperkte redactionele controle (correleert zowel met weinig als met veel interactie: monologen, Usenet) en ruime redactionele controle (correleert met weinig interactie: krantenmateriaal):

- Er is weinig of geen redactionele controle: de taalgebruiker heeft door de aard van de communicatie niet de mogelijkheid om het resultaat van de taalproductie aan te passen.
- Er is een beperkte redactionele controle: de taalgebruiker heeft de mogelijkheid om het resultaat van de taalproductie aan te passen.
- Er is een ruime redactionele controle: de taalgebruiker heeft niet alleen de mogelijkheid om het resultaat van de taalproductie aan te passen, hij/zij wordt in dat redactieproces ook geholpen door zijn/haar ervaring als professioneel taalgebruiker; bovendien wordt de individuele redactionele controle nadien ook nog eens herhaald door onafhankelijke professionele taalgebruikers (de eindredactie).

Als we, net als hierboven, even aannemen dat de mate van redactionele controle samenhangt met de graad van formaliteit (hoe hoger de redactionele controle, hoe hoger de formaliteit), dan kunnen we in deze paragraaf [HYP 3.6] testen:

[HYP 3.6] In formeel taalgebruik komt de rode volgorde [AUX+PART] vaker voor dan in informeel taalgebruik.

Tabel 3.9: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per mate van redactionele controle.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Geen redactionele controle	59.2% (1824/3081)	40.8% (1257/3081)
Weinig redactionele controle	44.9% (1174/2617)	55.1% (1443/2617)
Ruime redactionele controle	28.3% (2711/9589)	71.7% (6878/9589)

Uit tabel 3.9 komt een duidelijk positief lineaire trend naar voren: hoe meer redactionele controle, hoe meer rode woordvolgordes [AUX+PART] (bevestiging van [HYP 3.6]). Wegens het ordinale karakter van de verklarende variabele (*mate van redactionele controle*) gebruiken we in plaats van de Pearson χ^2 (die geen rekening houdt met een eventuele lineaire associatie) de krachtigere Mantel-Haenszel χ^2 (MH- χ^2) als testgrootte om de onafhankelijkheid van de verklarende en de responsvariabele te testen. De Mantel-Haenszel χ^2 laat zien dat er een significant lineaire associatie is (MH- $\chi^2 = 1028.38$, $df = 1$, $p < .0001$). Om de sterkte van de lineaire associatie te bepalen, wordt—eveneens wegens het ordinale karakter van de variabelen—de gammacoëfficiënt (γ) aangewend en niet de lambdacoëfficiënt.¹⁶ De gammacoëfficiënt kan een waarde tussen -1 en 1 aannemen: 1 staat daarbij voor een perfect positieve lineaire associatie, -1 voor een perfect negatieve lineaire associatie, 0 betekent dat er geen associatie is. De gammacoëfficiënt voor de associatie tussen de mate van redactionele controle en woordvolgorde geeft aan dat er een sterke, significante positief lineaire associatie is ($\gamma = 0.45$). Om na te gaan hoe representatief die steekproefgammacoëfficiënt is voor de (onbekende) populatiegammacoëfficiënt bakenen we een 95%-betrouwbaarheidsinterval af rond de steekproefgammacoëfficiënt. Dat doen we als volgt: B.I. = $\gamma \pm 1.96 * ASE$ (ASE staat voor *Asymptotic Standard Error*, asymptotische standaardfout). Voor de associatie tussen mate van redactionele controle en woordvolgorde heeft de ASE een waarde van 0.01, dus: $0.45 \pm 1.96 * 0.01 = [0.43 ; 0.46]$. Dat betekent met andere woorden dat de gammacoëfficiënt voor de volledige populatie een waarde aanneemt die zich met 95% zekerheid tussen 0.43 en 0.46 bevindt. Aangezien de waarde 0 (i.e. de associatie is niet significant) niet in het betrouwbaarheidsinterval ligt, kunnen we besluiten dat er een significant positief lineaire associatie is tussen de mate van redactionele controle en de woordvolgorde. Uit de CMH-analyses blijkt dat die associatie stabiel is over de verschillende strata van de controlevariabelen heen: *hulpwerkwoord* (*zijn* vs. *hebben* vs. *worden*; CMH- $\chi^2 = 1107.84$, $df = 1$, $p < .0001$), *geslacht* (m vs. v; CMH- $\chi^2 = 238.25$, $df = 1$, $p < .0001$) of *leeftijdscategorie* (> 44 vs. < 45; CMH- $\chi^2 = 224.81$, $df = 1$, $p < .0001$)¹⁷; er zijn geen redundantie- of interactie-effecten.

Ook die resultaten kunnen opnieuw geïnterpreteerd worden in termen van productiedruk: een dalende productiedruk zorgt voor een (potentiële) stijging van de redactionele controle op het eindresultaat. Een stijging van de redactionele controle lijkt dan samen te gaan met een stijging van het gebruik van de rode woordvolgorde [AUX+PART]. Als die redactionele controle dan ook nog eens uitgevoerd wordt door een min of meer professionele taalgebruiker, dan lijkt de rode volgorde [AUX+PART] in nog hevigere mate gekozen te worden.

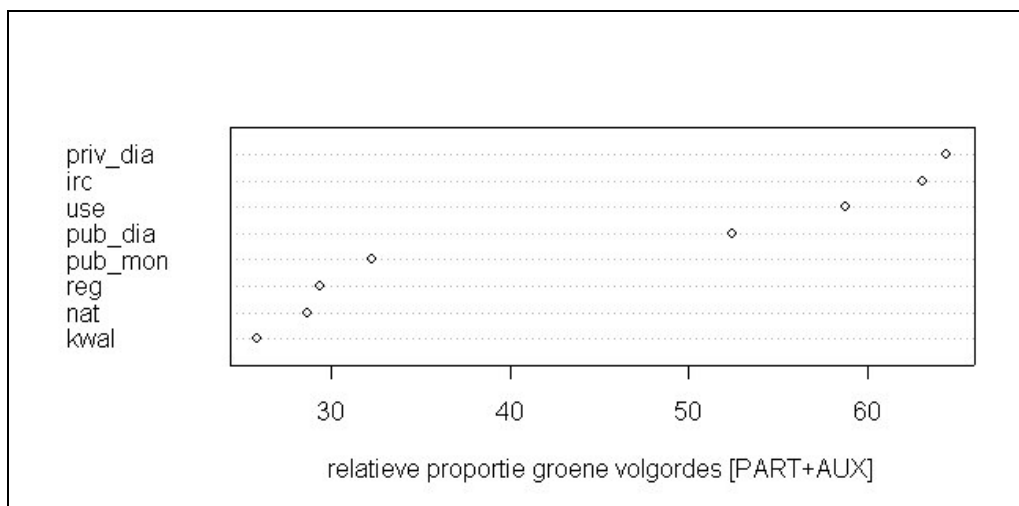
3.3.3.4 Modus, interactie en redactionele controle vergeleken en geherinterpreteerd

Uit het voorgaande overzicht is gebleken dat de drie bestudeerde dimensies een significant effect hebben op de keuze voor een van de beide syntactische varianten (in tegenstelling tot wat Zuckerman 2001 beweert: zie paragraaf 3.3.1). Het verschil in verklarende kracht tussen de tweede en derde dimensie is echter klein. Zoals hierboven al gezegd werd, is de derde dimensie, mate van redactionele controle, een verfijning¹⁸ van de tweede, mate van interactie, waardoor de communicatiesituaties met een ruime redactionele controle altijd samenvallen met de eenrichtingscommunicatie en de communicatiesituaties met weinig redactionele controle altijd samenvallen met de tweerichtingscommunicatie. Dat blijkt ook uit de kappatest (κ), waarmee we de mate van overeenkomst tussen de twee dimensies statistisch kunnen verifiëren (de categorieën met geen redactionele controle en beperkte redactionele controle zijn voor deze analyse samengenomen): $\kappa = 0.78$ (B.I. = 0.58-0.98). Aangezien κ een waarde kan aannemen die kleiner is dan of gelijk is aan 1, waarbij 1 staat voor een perfecte overeenkomst, kunnen we op basis van een kappawaarde van 0.78 besluiten dat de overeenkomst tussen de twee dimensies erg groot is. We zullen in het verdere verloop van dit artikel de tweede en derde dimensie dan ook niet meer van elkaar onderscheiden.

Verder hebben we bij de bespreking van de tweede dimensie, de mate van interactie, opgemerkt dat de keuze voor woordvolgorde niet zozeer bepaald leek te worden door wat in de eerste dimensie bestudeerd werd, de modus waarin de communicatie plaatsvindt, maar wel door wat we de ‘algemene productiedruk’ genoemd hebben. Dat vermoeden wordt bevestigd door figuur 3.1, waar het relatieve aandeel groene woordvolgordes [PART+AUX] over de verschillende subcomponenten

van het CONDIV- en CGN-corpus (R5) grafisch weergegeven wordt. De subcomponenten zijn op de verticale as geordend volgens stijgend aandeel groene woordvolgordes [PART+AUX]. De horizontale as geeft het precieze aandeel in termen van percentages aan.

Figuur 3.1: Relatieve proportie groene woordvolgordes [PART+AUX] in de verschillende CONDIV- en CGN-subcomponenten.



Figuur 3.1 laat twee duidelijk onderscheiden groepen zien: een eerste groep, die zich aan de linkerkant van de figuur bevindt, met een dominante voorkeur voor de rode volgorde [AUX+PART]; een tweede groep, die zich aan de rechterkant van de figuur bevindt, met een preferentie voor de groene volgorde [PART+AUX]. De eerste groep bevat de corpuscomponenten kwaliteitskranten (kwal), nationale (nat) en regionale populaire kranten (reg) en publieke monologen (pub_mon), de tweede groep bevat de publieke dialogen (pub_dia), Usenet (use), IRC (irc) en privédialogen (priv_dia). Op basis van die verdeling kunnen we alvast besluiten dat, conform de algemene bevindingen van Biber (1988), gesproken taal zich niet zo homogeen gedraagt als intuïtief gedacht wordt—de proportie groene woordvolgordes [PART+AUX] in het gesproken taalgebruik varieert van iets meer dan 30% naar bijna 70%—en dat gesproken en geschreven taal geen mooi afbakenbare gehelen zijn: publieke monologen vertonen net als het geschreven krantenmateriaal een frequenter rood-gebruik [AUX+PART], terwijl geschreven internetcommunicatie, net als het gesproken dialogische materiaal, neigt naar een dominant groen-gebruik [PART+AUX]. Het grote aandeel van rode woordvolgordes [AUX+PART] in de publieke monologen verklaart overigens ook waarom in tabel 3.7, waar het effect van

modus op woordvolgorde werd bestudeerd, nog ongeveer 50% rode woordvolgordes [AUX+PART] voorkwamen in het gesproken materiaal.

Op basis van figuur 3.1 kunnen we ook besluiten dat het niet de modus van de communicatieve situatie is die de corpuscomponenten binnen één groep met elkaar verbindt, maar wel de mate van interactie en redactionele controle: de ‘groene’ groep [PART+AUX] bevat alle corpuscomponenten waar er veel interactie en weinig redactionele controle aanwezig is, de rode groep [AUX+PART] bevat alle corpuscomponenten waar weinig interactie en middelmatig tot veel redactionele controle voorhanden is.¹⁹ Die hypothese wordt bevestigd door een conditionele analyse waarin het effect van modus op woordvolgorde geheranalyseerd wordt voor de verschillende maten van interactie apart. Op die manier zien we dat het significante effect van modus op woordvolgorde in de globale, samenvattende CMH-statistieken verdwijnt ($CMH-\chi^2 = 1.75$, $df = 1$, $p > .05$), hetgeen erop wijst dat het in tabel 3.7 vastgestelde effect van modus op woordvolgorde in feite een schijneffect is en dat een onderliggende variabele, de mate van interactie, voor het eigenlijke effect zorgt. De realiteit is echter nog iets genuanceerder. Het resultaat van de Breslow-Day-test wijst er namelijk op dat de o.r.’s voor de partiële tabellen niet homogeen zijn ($Br-D-\chi^2 = 9.18$, $df = 1$, $p = .002$). In concreto betekent dat dat er een interactie-effect is tussen eenrichtingscommunicatie en modus ($\chi^2 = 9.6$, $df = 1$, $p = .001$; o.r. = 0.83, B.I. = 0.73-0.93), het effect van modus op woordvolgorde verdwijnt in gevallen van tweerichtingscommunicatie ($\chi^2 = 1.4$, $df = 1$, $p > .05$). De dimensie *modus* die het verschil tussen gesproken en geschreven taal overkoepelt, bepaalt met andere woorden enkel de keuze voor woordvolgorde in niet-interactieve communicatiesituaties. In feite komt dat erop neer dat er een verschil bestaat tussen publieke monologen en krantenmateriaal (eenrichtingscommunicatie), maar niet tussen het internetmateriaal en het dialogische materiaal (tweerichtingscommunicatie).

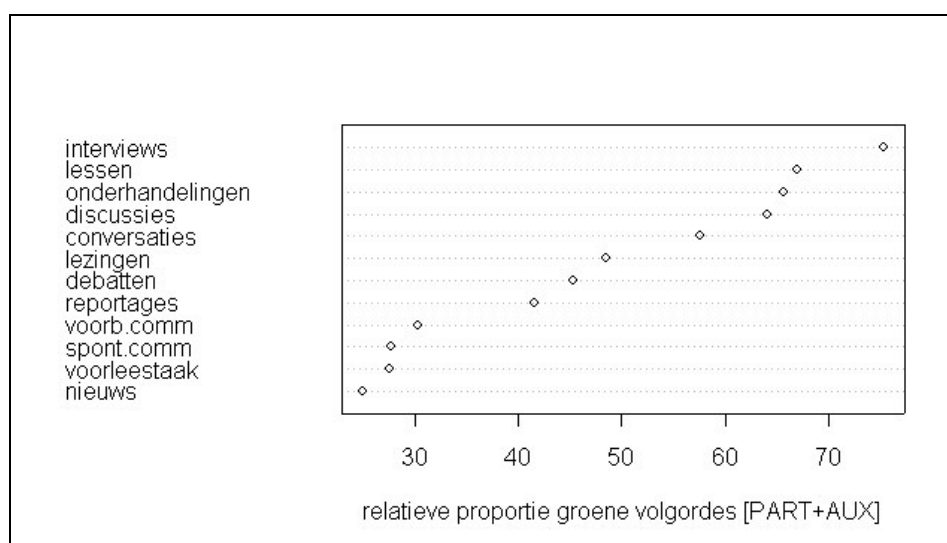
Die driedeling in krantenmateriaal, monologen en gesproken en geschreven dialogisch materiaal kan nu geïnterpreteerd worden in termen van stijgende on line productiedruk, met name respectievelijk afwezigheid van on line productiedruk (het resultaat van het productieproces kan aangepast worden door de auteur en de eindredactie), aanwezigheid van articulatoirische productiedruk (de grammaticale encoding van de geïntendeerde boodschap kan voorbereid en aangepast worden, de

fonetische en prosodische realisatie is echter gebonden aan de productie-eisen van een on line communicatieve situatie) en aanwezigheid van zowel articulatorische productiedruk als grammaticale encoderingsdruk.²⁰ Die gradatie van minder naar meer productiedruk is, zoals verwacht, negatief gecorreleerd met het voorkomen van de rode woordvolgorde [AUX+PART] ($\gamma = -0.48$, ASE = 0.01). Het lijkt er dus sterk op dat taalgebruikers de groene volgorde [PART+AUX] vervangen door de rode volgorde [AUX+PART] in omstandigheden waar de productiedruk laag of afwezig is en er daardoor tijd voorhanden is om het uiteindelijke product te herwerken en te onderwerpen aan een al dan niet bewust proces van controle.

De vraag die men zich daarbij kan stellen, is of die verhoogde aandacht voor het resultaat van de taalproductie een bewuste keuze voor de rode volgorde [AUX+PART] impliceert—is er sprake van een bewuste en geïntendeerde keuze voor de rode volgorde [AUX+PART]?— dan wel of het zo is dat de verhoogde aandacht voor de taalproductie de intentie om meer rode volgordes [AUX+PART] te produceren stimuleert, maar dat er geen sprake is van een bewuste intentie (zie Van Bree 2000 voor een bespreking van het verschil tussen intentie en bewustzijn). Voor de eerste optie pleiten de in de literatuur (zie Haeseryn 1990: 40-42 voor een overzicht) aangehaalde anti-Duits- of anti-dialectreflex (al is het mogelijk dat die ook onbewust werken) en de voorkeur van taalgebruikers om via variatie monotonie te vermijden: “Soms geeft de keus tussen beide woordschikkingen gelegenheid om een onaangename herhaling te vermijden” (Den Hertog 1972/1973). Daartegen kan echter ingebracht worden dat syntactische fenomenen over het algemeen aan de aandacht en de bewuste controle van taalgebruikers ontsnappen (Gerritsen 1991: 11, 20-21, Van Bree 2000: 24). Haeseryn (1990: 346) onderschrijft die laatste visie na een vergelijking van gedrag en oordeel over de volgordemogelijkheden in de werkwoordelijke eindgroep. Hij komt tot de conclusie dat het gedrag van de informanten geleid wordt door een onbewuste intentie om voor een bepaalde volgorde te kiezen: “We interpretern dit als een aanwijzing dat we hier niet te maken hebben met zomaar ideeën over wat de beste variant is, maar met echt geïnternaliseerde normen, die richtinggevend zijn voor het eigen gedrag”. Welke volgorde dat is, hoeft echter niet voor alle lagen van de bevolking dezelfde te zijn (zie bv. Haeseryn 1990: 185).

Dat de dimensies interactie en redactionele controle in termen van productiedruk geïnterpreteerd zijn, hoeft overigens niet te betekenen dat er geen andere interpretaties mogelijk zijn. Zo kunnen de communicatieve omstandigheden waarin de rode volgorde [AUX+PART] vaker voorkomt ook geïnterpreteerd worden in termen van journalistiek taalgebruik. Door het type materiaal dat in de beide corpora zit, kunnen we het krantenmateriaal niet vergelijken met materiaal dat in dezelfde modus geproduceerd is en dezelfde mate van interactie en redactionele controle heeft, zoals bijvoorbeeld literair materiaal. Als namelijk zou blijken dat er in literair materiaal helemaal geen voorkeur is voor de rode volgorde [AUX+PART], dan moeten we onze interpretatie op zijn minst bijstellen. In twee onderzoeken heeft De Schutter (1964, 1976) onderzoek gedaan naar dit type van syntactische variatie in literaire teksten. Daaruit blijkt dat 56.6% en 52.5% groene volgordes [PART+AUX] gekozen worden.²¹ In onze figuur zouden die ergens tussen de publieke dialogen en het Usenetmateriaal komen. Het lijkt er dus op dat de dominante preferentie voor de rode volgorde [AUX+PART] in ons krantenmateriaal en de publieke monologen ook geïnterpreteerd kan worden als een journalistieke trek. Dat die interpretatie niet zo gek is, kan ook aangetoond worden aan de hand van figuur 3.2, waar we het gesproken materiaal opgesplitst hebben volgens de kleinste eenheden die het CGN (R5) onderscheidt.

Figuur 3.2: Relatieve proportie groene woordvolgordes [PART+AUX] in verschillende communicatieve situaties (per CGN-component).



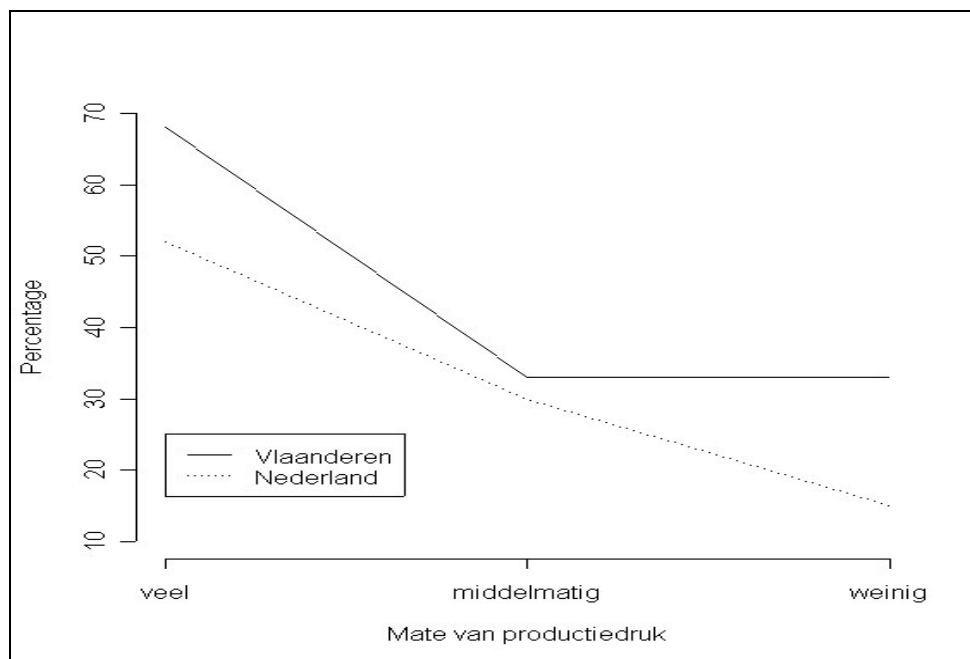
Opvallend is dat vier van de vijf communicatieve situaties waar het vaakst de rode volgorde [AUX+PART] gehanteerd wordt als *journalistiek* kunnen worden gecategoriseerd (nieuws, spontane commentaar, voorbereide commentaar en reportages). Dat hoeft uiteraard niet te betekenen dat de interpretatie in termen van productiedruk onhoudbaar is. Het kan namelijk zo zijn dat door de hoge mate van productiedruk taalgebruikers terugvallen op de groene woordvolgorde [PART+AUX], terwijl ze bij niet-interactieve communicatie kunnen kiezen tussen de groene [PART+AUX] (literair taalgebruik) en de rode volgorde [AUX+PART] (journalistiek taalgebruik).

Hoe het ook moge zijn, het is duidelijk dat de resultaten en interpretaties die in deze sectie voorgesteld zijn in vervolgonderzoek kritisch getoetst zullen moeten worden. In eerste instantie denken we daarbij dan aan een analyse van het gebruik van de beide volgordevarianten in verschillende types literair materiaal. Daaruit moet blijken in hoeverre de hier verkregen resultaten uitsluitend te danken zijn aan het journalistieke taalgebruik dan wel algemeen geldig zijn voor het redactioneel gecontroleerde en niet-interactieve taalgebruik. Dat onderzoek naar literair taalgebruik zal dan in ieder geval wel gecontroleerd moeten worden voor verschillende literaire stromingen, de datum van publicatie, het type spreker dat de werkwoordelijke eindgroep produceert, de context waarin de eindgroep voorkomt, eventueel zelfs voor een aantal persoonlijke gegevens van de auteur, en dergelijke meer. In tweede instantie moet de procesmatige hypothese in een goed gecontroleerde experimentele context getest worden om een antwoord te kunnen bieden op de vraag of de taalproductieomstandigheden een invloed hebben op het uiteindelijke product, i.c. de volgorde in de tweeledige werkwoordelijke eindgroep.

3.4 De interactie tussen het regio- en het registereffect

De vraag die men zich nu nog kan stellen, is hoe het geobserveerde regionale effect in verband gebracht kan worden met het registereffect. Om die vraag te beantwoorden, werd in figuur 3.3 de proportie groene volgordes [PART+AUX] uitgesplitst per regio (Nederland: puntlijn versus Vlaanderen: volle lijn) en per mate van productiedruk (veel productiedruk: gesproken en geschreven dialogen, middelmatige productiedruk: monologen, weinig productiedruk: krantenmateriaal):

Figuur 3.3: Relatieve proportie groene woordvolgordes [PART+AUX] per mate van productiedruk en per regio.



Figuur 3.3 toont dat over de verschillende graden van productiedruk heen een duidelijk regionaal verschil behouden blijft, met uitzondering van de communicatieomstandigheden waar sprake is van een middelmatige productiedruk: daar werd geen statistisch significant verschil vastgesteld tussen het Belgische en het Nederlandse Nederlands.²² Dat wijst erop dat zowel Nederlanders als Vlamingen in monologische communicatiesituaties dezelfde woordvolgordepreferenties hebben. Daarnaast blijkt ook dat er binnen elke regio een duidelijk lineaire trend is tussen mate van productiedruk en keuze voor de rode woordvolgorde [AUX+PART].²³ Die lineaire trend is echter duidelijker voor het Nederlandse Nederlands dan voor het Belgische Nederlands waar de voorkeur voor de groene volgorde [PART+AUX] niet verder daalt met een dalende productiedruk. Ondanks die kanttekeningen kunnen we hieruit toch besluiten dat zowel de regio als het register (in termen van productiedruk) een rol spelen bij de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde. In concreto komt het erop neer dat zowel Vlamingen als Nederlanders evenzeer de neiging hebben om op de groene volgorde [PART+AUX] terug te vallen bij een stijgende productiedruk, maar dat de mate waarin teruggevallen wordt op de groene volgorde [PART+AUX] verschilt per regio: in het Belgische Nederlands wordt de groene volgorde [PART+AUX] vaker gebruikt dan in het Nederlandse Nederlands.

3.5 Samenvatting

In dit hoofdstuk hebben we de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep met een deelwoord en een hulpwerkwoord vanuit een contextueel perspectief bestudeerd. De analyses waarover in dit hoofdstuk gerapporteerd is, hebben daarbij duidelijk gemaakt dat de keuze voor de groene [PART+AUX] of rode [AUX+PART] volgorde beïnvloed wordt door de *regio* en het *register* waarin de taalproductie plaatsvindt. We vatten voor de beide factoren de belangrijkste bevindingen samen.

3.5.1 Regio

Ondanks een algemene voorkeur voor de rode volgorde [AUX+PART] in onze dataset maken de resultaten van het macroperspectivische onderzoek naar de regionale verschillen in het gebruik van de groene [PART+AUX] en rode [AUX+PART] volgorde toch duidelijk dat taalgebruikers in Vlaanderen vaker gebruik maken van de groene volgorde [PART+AUX] dan taalgebruikers in Nederland. Die resultaten zijn grosso modo in overeenstemming met wat bekend was uit vroeger onderzoek. Vervolgens werd ook nagegaan of het vastgestelde effect van regio niet terug te voeren is op een verborgen, onderliggende variabele, zoals het type hulpwerkwoord, het geslacht of de leeftijd. Dat bleek niet het geval te zijn, zodat geconcludeerd kan worden dat het regionale verschil in de keuze voor een van beide woordvolgordevarianten een reëel verschil is. Het verschil in voorkeur tussen de beide delen van het Nederlandse taalgebied werd voorlopig geïnterpreteerd in termen van een ander standaardiseringstraject: redenerend vanuit datgene wat geweten is voor het lexicon (Geeraerts et al. 1999), werd aangenomen dat het voorkeursverschil geleidelijk aan slinkt, in die zin dat het Belgische Nederlands naar het Nederlandse Nederlands toegroeit (het gebruikt gaandeweg vaker de rode volgorde [AUX+PART]). Uit het diachrone onderzoek van Coussé (2003) lijkt er alvast een opmars te zijn van rode woordvolgordes [AUX+PART] in de tijdspanne van de 13^{de} tot de 19^{de} eeuw in Hollandse teksten. De vraag is of die opmars zich ook in de 20^{ste} eeuw heeft doorgezet, of Vlaanderen die tendens is gevolgd (vgl. de grotere voorkeur voor de rode volgorde [AUX+PART] in Vlaams-Brabant en Antwerpen ten opzichte van de andere Vlaamse provincies; paragraaf 3.2.3) en of de eventuele kloof tussen het

Belgische en Nederlandse Nederlands in volgordevoorkeur afgenomen is. Een andere optie is uiteraard dat in de beide regio's anders gebruik gemaakt wordt van de syntactische variatie (bv. pragmatisch-functioneel). Voorlopig zijn dat uiteraard niet meer dan suggesties voor verder onderzoek.

Vanuit een microperspectivische benadering van de woordvolgordevariatie konden we vaststellen dat het Belgische Nederlands, ondanks een algemene voorkeur voor de groene volgorde [PART+AUX], geen homogeen geheel is: het Brabantse materiaal bevat namelijk vaker dan gemiddeld rode volgordes [AUX+PART] en sluit daarmee aan bij wat gebruikelijk is in het Nederlandse Nederlands. Het Oost-Vlaamse materiaal daarentegen bevat vaker groene volgordes [PART+AUX]. Voor Nederland konden we niets terugvinden van wat uit vroeger dialectologisch onderzoek gebleken was, met name de extreme voorkeur voor de groene volgorde [PART+AUX] in de noordelijke provincies en de voorkeur voor de rode volgorde [AUX+PART] in Nederlands Limburg. Dat betekent echter niet dat de resultaten van dat vroeger onderzoek hiermee weerlegd zijn, aangezien (i) we niet over een corpus van dialectologisch materiaal beschikten en (ii) de absolute frequenties in sommige gevallen niet erg hoog waren (hetgeen vermoedelijk ook de extreme positie van het Zeeuwse materiaal—73.1% rode volgordes [AUX+PART], veruit de grootste proportie voor het hele Nederlandse taalgebied—in onze dataset verklaart).

3.5.2 Register

Register werd in deze studie, in tegenstelling tot de dichotome classificaties in eerder onderzoek, via drie dimensies geoperationaliseerd: modus, mate van interactie en mate van redactionele controle.

- *Modus* – Geheel in overeenstemming met de resultaten van eerder onderzoek kon vastgesteld worden dat er een significant verschil is tussen gesproken en geschreven taal: taalgebruikers blijken de groene volgorde [PART+AUX] vaker te gebruiken in de gesproken taal dan in de geschreven taal. Die vaststelling werd, conform met de bevindingen van Biber (1988: 113-114), geïnterpreteerd tegen de achtergrond van de specifieke eigenschappen van het productieproces in gesproken versus geschreven taal (planning van de uiting).

- *Mate van interactie* – Werkwoordelijke eindgroepen die geproduceerd worden in omstandigheden waar twee of meer gesprekspartners met elkaar interageren (gesproken dialogen, geschreven dialogen op babbelkanalen), vertonen significant vaker groene volgordes [PART+AUX] dan in monologische situaties (gesproken monologen, krantenartikelen).
- *Mate van redactionele controle* – Werkwoordelijke eindgroepen die geproduceerd zijn in omstandigheden waar de taalgebruiker door de aard van de communicatie niet de mogelijkheid heeft om het resultaat van de taalproductie aan te passen vertonen vaker groene volgordes [PART+AUX] dan in situaties waar het taalgebruik beperkt gecontroleerd wordt; taaluitingen die ruim gecontroleerd worden vertonen op hun beurt dan weer minder groene volgordes [PART+AUX] dan beperkt gecontroleerde uitingen. Met andere woorden: hoe groter de mate van redactionele controle, hoe groter de voorkeur voor de rode woordvolgorde [AUX+PART].

De effecten van de drie registerdimensies op de woordvolgorde laten zien dat taalgebruikers in omstandigheden waar de productiedruk groot is—i.e. er zijn tijdsbeperkingen op de fonetische, prosodische en/of algehele encoding van de geïntendeerde boodschap—vaker voor de groene volgorde [PART+AUX] kiezen. Omgekeerd betekent dat dan dat het erop lijkt dat de groene volgorde [PART+AUX] vervangen wordt door de rode volgorde [AUX+PART] in omstandigheden waar de productiedruk laag of afwezig is en er daardoor tijd voorhanden is om het uiteindelijke product te herwerken en te onderwerpen aan een al dan niet bewust proces van controle. Bij een onderlinge vergelijking van de drie dimensies bleek dat de tweede en de derde dimensie (interactiviteit en redactionele controle) bijna identiek hetzelfde doen en dus samengenomen kunnen worden, dat gesproken taal, in tegenstelling tot wat intuïtief vaak aangenomen wordt, helemaal geen homogeen gedrag vertoont en dat gesproken en geschreven taal geen mooi afbakenbare gehelen zijn. Er werd verder nog vastgesteld dat het verschil tussen gesproken en geschreven taal gedeeltelijk teruggevoerd kan worden op een verschil in interactiviteit/redactionele controle, zodat de oorspronkelijke bestudering van register langs drie dimensies in feit teruggevoerd kan worden op het verschil tussen krantenmateriaal, gesproken monologen en gesproken en geschreven dialogen. Tot slot werd er nog op gewezen dat de interpretatie in termen van productiedruk hoegenaamd niet hoeft te betekenen dat er

geen andere interpretaties meer mogelijk zijn (zoals een interpretatie in termen van journalistiek versus niet-journalistiek taalgebruik).

Noten:

1. We kunnen hier uiteraard enkel uitspraken doen over de proporties in de hele populatie voor zover we aannemen dat de stratificatie in de hele populatie identiek is aan de stratificatie in de steekproef (bv. wanneer we aannemen dat het aandeel gesproken materiaal binnen onze steekproef identiek is aan het aandeel gesproken materiaal in de volledige populatie).
2. We baseren ons overzicht op Frings & Vandenheuvel 1921—die zinnen 8, 18 en 20 uit de vragenlijst van Wenker aan Vlaamse informanten hebben voorgelegd—Van den Bergh 1949, Meeussen & Vanacker 1951, Pauwels 1953, Stroop 1970 en Gerritsen 1991.
3. Dat een dergelijk effect ook in de (gesproken en geschreven) productie van Nederlandse tweeledige werkwoordelijke eindgroepen speelt, heeft het onderzoek van Hartsuiker & Westenberg (2000) aangetoond (zie hoofdstuk 6).
4. In tabel 3.3 zal blijken dat de mate van meegaandheid hoog is. In dat opzicht is het opmerkelijk dat Haeseryn de resultaten van die tests toch nog gebruikt om de invloed van allerlei factoren op de keuze voor de woordvolgorde na te gaan.
5. Net zoals voor proporties wordt ook voor o.r.-waarden een betrouwbaarheidsinterval berekend. Dat geeft de grenzen aan waarbinnen de o.r. in de volledige populatie gesitueerd moet worden.
6. Een andere (deels gerelateerde) hypothese voor het verschil in voorkeur is dat de Nederlandse taalgebruikers gefixeerd zijn op de rode volgorde vanuit een anti-Duitse reflex (Stroobants 1997: 148).
7. Die vaststelling zegt echter niets over de reële verdeling van werkwoordelijke eindgroepen over verschillende sociale klassen, maar wel over het nauwelijks voorhanden zijn van gegevens voor de lage sociale klassen (opleiding en beroep) in CGN (R5). Zo zijn er bijvoorbeeld slechts 10 laagopgeleiden (op een totaal van 1560 taalgebruikers) voor het hele Nederlandse taalgebied in de vijfde release van het CGN beschikbaar. Voorlopig kunnen dan ook geen representatieve uitspraken gedaan worden over de verdeling van de woordvolgordevarianten over verschillende sociale klassen.
8. Dat net bij het hulpwerkwoord van het passief *worden* een gebrek aan data vastgesteld wordt, hoeft ons niet te verbazen. Passieve constructies worden gebruikt om de patiëns meer aandacht te geven: “Discourse with very frequent passive constructions is typically abstract and technical in content, and formal in style” (Biber 1988: 112). Net die kenmerken conflicteren met het materiaal dat hier bestudeerd wordt, met name spontaan tot stand gekomen, niet-formele, niet-publieke dialogen.
9. Wegens de lage frequenties konden geen betrouwbare conditionele analyses uitgevoerd worden. Een reductie van het aantal provincies was geen mogelijkheid, omdat we dan bv. de extreme regio

Zeeland samen zouden moeten nemen met minstens een andere regio, terwijl de conditionele analyses nu net tot doel hadden om te kijken of het extreme gedrag van de Zeeuwen geen schijneffect is.

10. In dit verband moet ook opgemerkt worden dat zowel het gesproken als het geschreven Vlaamse materiaal veel meer werkwoordelijke eindgroepen bevat dan het Nederlandse materiaal, ondanks het feit dat het Nederlands-Nederlandse corpusmateriaal meer tokens in de gesproken en informeel geschreven componenten bevat dan het Belgische. Voor het CGN (R5) konden we daarnaast ook nagaan dat het hoge aantal eindgroepen in het Vlaamse materiaal samenhangt met een verhoogd voorkomen van het aantal bijzinnen. Die constatering verleidt ons ertoe de hypothese te formuleren, gebaseerd op de bevindingen van Biber, dat de Vlaamse taalgebruiker—in tegenstelling tot de Nederlandse—ofwel persoonlijker is ofwel minder vlot de taal hanteert (met name door meer constructies met een lage lexicale densiteit te gebruiken).
11. Die resultaten bevestigen ook ons vermoeden uit paragraaf 3.1 dat de dominantie van de rode volgorde [AUX+PART] in onze dataset teruggevoerd kan worden op de proportionele dominantie van geschreven materiaal in de dataset.
12. We kunnen hier niet controleren voor sekse of leeftijd, omdat de opsplitsing in geschreven en gesproken taalgebruik samenvalt met het verschil tussen het CONDIV-corpus en het CGN (R5) en we enkel voor CGN (R5) informatie hebben over het geslacht en de leeftijd van de taalgebruiker.
13. Dat is de laatste stap in een complex en incrementeel encoderingsproces: de geïntendeerde boodschap wordt in eerste instantie grammaticaal geëncodeerd, i.e. de geschikte woorden worden uit het mentale lexicon gehaald, de woorden wordt een functie toegekend en vervolgens wordt de positie van de woorden ten opzichte van elkaar bepaald. In tweede instantie volgt dan het fonologisch encoderingsproces dat bestaat uit de articulatie van de fonologische structuur van een uiting in termen van fonologische segmenten en prosodie van grotere gehelen (Bock & Levelt 1994: 946).
14. Met uitzondering van de CMH-resultaten voor de controlevariabele *hulpwerkwoord* gelden de CMH-resultaten uiteraard enkel voor de CGN-data, aangezien we voor CONDIV niet over leeftijds- en geslachtsinformatie beschikken.
15. In tegenstelling tot wat de term *redactioneel* suggereert, wordt de term hier gebruikt om te refereren aan situaties van ‘externe’ controle (door een redactiechef, de eindredactie, een collega,...) en ‘interne’ controle (door de auteur zelf, *self-monitoring*).
16. De lambda-coëfficiënt (λ), die verderop nog aan bod komt, is een associatiemaat voor categorische variabelen die als alternatief voor de o.r. aangewend kan worden in het geval dat een of meer van de variabelen in de bivariate analyse meer dan twee waarden bevat (dan kan de o.r. niet berekend worden). De waarden van de lambdamaat variëren tussen 0 en 1, waarbij 0 staat voor geen associatie en 1 voor een sterke associatie.
17. Met uitzondering van de CMH-resultaten voor de controlevariabele *hulpwerkwoord* gelden de CMH-resultaten uiteraard enkel voor de CGN-data, aangezien we voor CONDIV niet over leeftijds- en geslachtsinformatie beschikken.

18. Al kan het uiteraard niet als een zuivere verfijning beschouwd worden, aangezien we een *cut off point* verliezen: de grens tussen de categorieën van de variabele *mate van interactie* heeft geen pendant bij de variabele *mate van redactionele controle*.
19. De interpretatie in termen van redactionele controle levert echter problemen op: hierboven hadden we zowel de monologen als het Usenetmateriaal gecategoriseerd als taalgebruik in communicatieve situaties met een beperkte redactionele controle. In dat opzicht is het merkwaardig dat de proporties groene volgordes [PART+AUX] voor de monologen en Usenet zover uit elkaar liggen.
20. Aangezien het verschil tussen gesproken en geschreven taalgebruik slechts gedeeltelijk relevant is, kan men zich hier de vraag stellen of de interpretatie in termen van articulatorische productiedruk terecht is. Immers, indien er sprake zou zijn van articulatorische productiedruk, dan zou men verwachten dat er tussen het gesproken dialogische materiaal en het internetmateriaal ook significante verschillen vastgesteld zouden moeten worden, quod non. Een alternatieve verklaring zou kunnen zijn dat het verschil tussen de gesproken monologen en het krantenmateriaal te wijten is aan nuances in de redactionele controle. Toekomstig onderzoek zal hier uitsluitsel over moeten geven.
21. De frequenties zijn overgenomen uit De Schutter (1964) en (1976) en hebben enkel betrekking op tweeledige werkwoordelijke eindgroepen met een deelwoord en *zijn*, *hebben* of *worden* als hulpwerkwoord (de frequenties van de constructies met andere hulpwerkwoorden zijn hier buiten beschouwing gelaten). De resultaten in De Schutter (1964) zijn gebaseerd op een corpus van Nederlands literair materiaal, de resultaten in De Schutter (1976) zijn gebaseerd op Vlaams en Nederlands literair materiaal. Aangezien de onderzoekscontext in De Schutter (1964, 1976) niet identiek is aan de onze (De Schutter bestudeert bv. eindgroepen in verschillende types bijzinnen), is de vergelijking hier louter indicatief bedoeld.
22. De volledige resultaten van de CMH-analyses met regio als verklarende variabele en mate van productiedruk als controlevariabele: er is een globaal significant effect van regio op de woordvolgorde, gecontroleerd voor register (CMH- $\chi^2 = 362.71$, $df = 1$, $p < .0001$; o.r. = 0.46, B.I. = 0.43-0.50). Op basis van het resultaat van de Breslow-Day-test wordt duidelijk dat de o.r. over de verschillende partiële tabellen heen niet homogeen is (Br-D- $\chi^2 = 46.60$, $df = 2$, $p < .0001$), wat wijst op een interactie-effect: het verschil tussen de beide regio's is significant in communicatiesituaties met veel productiedruk ($\chi^2 = 112.51$, $df = 1$, $p < .0001$; o.r. = 0.51, B.I. = 0.45-0.58) en weinig productiedruk ($\chi^2 = 284.03$, $df = 1$, $p < .0001$; o.r. = 0.37, B.I. = 0.33-0.42), maar niet voor communicatiesituaties met een middelmatige productiedruk ($\chi^2 = 0.97$, $df = 1$, $p > .05$).
23. De volledige resultaten van de CMH-analyses met mate van productiedruk als verklarende variabele en regio als controlevariabele: er is een globaal significant effect van mate van productiedruk op de woordvolgorde, gecontroleerd voor regio (CMH- $\chi^2 = 1499.61$, $df = 1$, $p < .0001$). Het verschil tussen de verschillende types register is zowel significant voor het Belgisch-

Nederlandse materiaal ($MH-\chi^2 = 800.49$, $df = 1$, $p < .0001$; $\gamma = 0.48$, $ASE = 0.02$) als voor het Nederlands-Nederlandse materiaal ($MH-\chi^2 = 728.77$, $df = 1$, $p < .0001$; $\gamma = 0.63$, $ASE = 0.02$).

Hoofdstuk 4

De prosodische dimensie

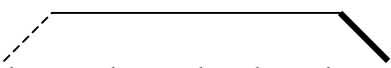
4.1 Inleiding

Langs een tweede dimensie hebben onderzoekers de woordvolgordevariatie in de werkwoordelijke eindgroep in verband gebracht met het streven naar een evenwichtige verdeling van accenten in de zin. Dat de accentuering een invloed zou kunnen uitoefenen op de woordvolgorde is iets wat al vroeg gesuggereerd werd (vgl. bv. Overdiep 1949: 530-531, Pauwels 1953: 102-105), overigens niet alleen voor Nederlandse werkwoordelijke eindgroepen, maar ook voor Duitse werkwoordelijke eindgroepen, die in vroegere taalstadia een soortgelijke variatie vertoonden (Ebert 1998), en voor andere syntactische fenomenen (zie bv. McDonald et al. 1993). De aantrekkelijkheid en plausibiliteit van een relatie tussen accentuering en woordvolgorde en de eensgezindheid onder wetenschappers staat echter in schril contrast met de vaagheid waarmee op de accentuering een beroep gedaan wordt. Illustratief daarvoor is het volgende citaat van Den Hertog (1972/1973: 44): “In de regel is het laten volgen van de persoonsvorm op deelwoord of infinitief vloeiender dan het omgekeerde. De meeste taalgebruikers zullen in dergelijke gevallen geen rekening houden met de volgorde; wie minder onverschillig is met betrekking tot de vorm waarin hij zijn gedachten uit, laat zich meestal door zijn gehoor leiden”. Wat Den Hertog nu precies ‘vloeiender’ vindt of wat het betekent om zich ‘door zijn gehoor te laten leiden’, is niet duidelijk. Naast het gebruik van dergelijke vage aanduidingen zijn ook andere termen gebruikt om naar de suprasegmentele invloed te verwijzen, zoals ritme, accent, klemtoon, melodie, intonatie en eufonie, maar nergens wordt heel precies gedefinieerd wat daar dan precies mee bedoeld is. Niet voor niets

zegt de ANS (1997: 1068) daarom in haar jongste editie dat het principe van de evenwichtige klemtoonverdeling vooralsnog niet exact te beschrijven is.

Rond diezelfde tijd echter gaf De Schutter (1996; zie ook Swerts 1998: 305-306) het onderzoek naar het verband tussen prosodie en woordvolgorde een nieuwe wending door het accentueringsprincipe concreter gestalte te geven. Zo beschouwt hij de variatie in de werkwoordelijke eindgroep als een relatief correctiemechanisme dat gebruikt kan worden indien een van de volgende prosodische principes geschonden dreigt te worden: het principe van het platte-hoedpatroon, dat een van de basisintonatiecontouren van het Nederlands is ('t Hart et al. 1990: 38-67) en het principe van de evenwichtige verdeling van zinsaccenten. Het eerste principe zegt dat een normale Nederlandse mededelende zin een toonhoogtestijging (stippellijn) en toonhoogtedaling (lijn in vetjes) kent op niet-adjacente syllabes (voorbeeld uit 't Hart et al. 1990: 116):

(6) We gaan vanavond in ieder geval naar de schouwburg.



Om de toonhoogtedaling te kunnen realiseren heeft een spreker wel een zekere prosodische ruimte nodig, i.e. een of meerdere onbeklemtoonde syllabes. Vandaar dat in zinnen met een werkwoordelijke eindgroep als laatste constituent de groene volgorde [PART+AUX], met het onbeklemtoonde hulpwerkwoord na het zinsaccentdragende deelwoord, de voorkeur wegdraagt. Het zinsaccentdragende deelwoord aan het eind van een zin (de rode volgorde [AUX+PART]) zou een minder ideale situatie opleveren, omdat dan aan het eind van de zin zowel prominentie verleend moet worden als een grens gemarkeerd. Dat betekent echter niet dat de rode volgorde [AUX+PART] in een dergelijke context volgens de intonatiegrammatica ongrammaticaal zou zijn, maar het is volgens De Schutter in ritmisch en melodisch opzicht minder geslaagd. De rode volgorde [AUX+PART] wordt volgens De Schutter echter wel geprefereerd—we zijn nu aanbeland bij het tweede principe—als een zinsaccent onmiddellijk aan de werkwoordelijke eindgroep voorafgaat en een accentenbotsing vermeden moet worden: door het hulpwerkwoord tussen het preverbaal zinsaccent en het zinsaccent op het deelwoord te plaatsen, wordt een te snelle opeenvolging van twee zware accenten vermeden.^{1, 2} Voor De Schutter worden twee geaccentueerde syllabes idealiter door 2 à 3 onbeklemtoonde syllabes van elkaar

gescheiden (een argument daarvoor geeft hij echter niet). Dat impliceert tevens dat de afstand tussen twee geaccentueerde syllabes ook niet te groot mag zijn: meer dan drie onbeklemtoonde lettergrepen na mekaar noemt De Schutter minder ‘gelukkig’.

Laten we de twee principes, waarnaar we in het vervolg zullen verwijzen als het *accentueringsprincipe* of de *accentueringshypothese*³, illustreren aan de hand van de volgende vier voorbeelden uit De Schutter (1996: 215) (- = geen accent, / = secundair accent, // = hoofdaccent):

(7a) (dat) we daar toen een onbekende gezien hebben.

- - - // - - / - -

(7b) (dat) er toen ineens een beeld verschenen is.

- // - / - -

(7c) (dat) ik mijn vriendje naar huis heb gebracht.

- // - - /

(7d) (dat) we daar toen een onbekende hebben gezien.

- - - // - - - - /

Het (a)-voorbeeld is conform het accentueringsprincipe: doordat de zin eindigt met twee zwakbetoonde lettergrepen is er voldoende prosodische ruimte om de toonhoogtedaling vanuit de tweede lettergreep van het deelwoord te realiseren en er is ook geen botsing van accenten. Het (b)-, (c)- en (d)-voorbeeld zijn minder ideaal doordat respectievelijk twee zinsaccenten te dicht op elkaar staan (*BEELD* en *verSCHEnen*), de toonhoogtedaling samen met de prominentieverlening moet plaatsvinden (*geBRACHT*) en het interval van zwak betoonde lettergrepen tussen *onbeKENde* en *geZIEN* te groot is.

Niettegenstaande De Schutter (1996) het onderzoek naar de samenhang tussen accentuering en woordvolgorde nieuw leven ingeblazen heeft door een aantal zaken die voordien onduidelijk gebleven waren concreet te maken, zijn er toch nog steeds een heel aantal zaken ongewis. Zo maakt De Schutter gebruik van de notie *zinsaccent* zonder er een definitie van te geven of zonder aan te geven hoe je in de praktijk een zinsaccent kan herkennen: is het een zeer specifieke akoestische notie, gecorreleerd met fonetische kenmerken als toonhoogte, amplitude en duur (Booij 1995: 96-124, Nooteboom 1997: 670) of is het een abstracte perceptuele notie, waarbij zinsaccenten als relatieve prominentieverleners beschouwd worden, rekening houdend met

prosodische, semantische, syntactische en/of pragmatische eigenschappen (Collier & Droste 1982: 34) of is het een puur syntactische of pragmatische notie? Daarnaast is ook onduidelijk hoe je een lage grenstoon kan herkennen.⁴ Door die definatorische problemen is het moeilijk om een aantal van De Schutters veronderstellingen te taxeren, zoals bijvoorbeeld de veronderstelling dat het hoofdwerkwoord en de constituent die de focus draagt (hetgeen in de praktijk kan samenvallen met het hoofdwerkwoord) steeds een zinsaccent krijgen en dat een hulpwerkwoord nooit een zinsaccent krijgt.

Het hoeft geen betoog dat dergelijke onduidelijkheden een directe operationalisering van de accentuering in het vervolg van dit hoofdstuk zullen bemoeilijken. De voorliggende studie zal dan ook niet voor een definitieve verheldering van het accentueringsprincipe kunnen zorgen, maar ze is er wel op gericht om een vooruitgang te bewerkstelligen door een aantal nieuwe, specifiekere tests uit te voeren waarin de impact van prosodische eigenschappen op de woordvolgorde in geschreven taal⁵ bestudeerd wordt. Daarbij zullen de veronderstellingen van De Schutter, tot nader order, overgenomen worden, i.e. het deelwoord en de gefocusseerde constituent krijgen een zinsaccent en een hulpwerkwoord krijgt nooit een zinsaccent.

In wat volgt zal het effect van de accentdistributie in drie stappen onderzocht worden: we beginnen met een studie naar de invloed van de accentverdeling binnen de eindgroep zelf (paragraaf 4.2) en analyseren vervolgens het effect van de accentverdeling in de linkercontext van de eindgroep (paragraaf 4.3) en het effect van de accentverdeling in de rechtercontext van de eindgroep (paragraaf 4.4).⁶ De factoren die daarbij aan bod zullen komen, zijn in tabel 4.1 opgesomd:

Tabel 4.1: Overzicht van de factoren in dit hoofdstuk.

Accentverdeling binnen de eindgroep	Accentverdeling in de linkercontext	Accentverdeling in de rechtercontext
Morfologische bouw van het deelwoord	Definietheid van de laatste preverbale substantiefgroep	Aan- vs. afwezigheid van een laatste zinsplaats
Afstand tussen het deelwoordaccent en het begin van de werkwoordelijke eindgroep	Informationaliteit van de laatste preverbale substantiefgroep	Grammaticale relatie van de laatste zinsplaats met het hoofd
Afstand tussen het deelwoordaccent en het einde van de werkwoordelijke eindgroep	Inherentie van de laatste preverbale constituent	Plaats van het accent op het eerste postverbale (inhouds)woord
	Lengte van het middenstuk	
	Plaats van het accent op het laatste preverbale woord	

Omdat het streven naar een evenwichtige accentverdeling in de zin geen rekening houdt met syntactische grenzen (linkercontext vs. eindgroep vs. rechtercontext), zullen we in paragraaf 4.5 het samenspel tussen de accentdistributie in de drie contexten bestuderen. Paragraaf 4.6 biedt een ruime samenvatting van de verwezenlijkingen in dit hoofdstuk.

Vooraleer echter het onderzoek naar de prosodische dimensie te beginnen, staan we eerst stil bij de steekproef die voor deze en de andere taalinterne dimensies (hoofdstuk 4-7) gebruikt zal worden. Zoals in hoofdstuk 2 aangekondigd werd, zal de studie naar de taalinterne krachten in het woordvolgordeonderzoek om puur praktische redenen gecontroleerd worden voor regionale effecten en registereffecten. Meer bepaald concentreren we ons op tweeledige werkwoordelijke eindgroepen uit de CONDIV-component *De Standaard supraregionaal*. De zoekopdracht in die component leverde 2390 observaties op die als volgt over de twee syntactische varianten verdeeld zijn:

Tabel 4.2: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde in de dataset.

Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
33.01%	66.99%
(789/2390)	(1601/2390)

Tegen de achtergrond van de onderzoeksresultaten in hoofdstuk 3 is het geen complete verrassing dat de rode volgorde [AUX+PART] in tabel 4.2 twee derde van de volledige dataset inneemt: het *De Standaard*-materiaal in de dataset kan nl. gekarakteriseerd worden als formeel geschreven Belgisch Nederlands zonder interactie en met een hoge mate van redactionele controle. Om de representativiteit van de steekproefproporties in tabel 4.1 na te gaan, bakenen we ook hier weer 95%-B.I. af. Die B.I. reiken van 31.16% tot 34.92% voor de groene volgorde [PART+AUX] en van 65.08% tot 68.84% voor de rode volgorde [AUX+PART], zodat we met 95% zekerheid kunnen zeggen dat de steekproefproporties in tabel 4.1 representatief zijn voor de volledige populatie en dat met andere woorden het proportionele verschil tussen de beide volgordes een reëel verschil is (dat ook hier weer statistisch significant is: $\chi^2 = 275.8762$, $df = 1$, $p < .0001$).

4.2 De impact van de accentverdeling in de werkwoordelijke eindgroep

In deze sectie gaan we na wat de invloed is van de accentverdeling in de werkwoordelijke eindgroep op de keuze voor de groene [PART+AUX] of rode [AUX+PART] woordvolgorde. Na een overzicht van de verwezenlijkingen van vroeger onderzoek (paragraaf 4.2.1) en de hypothesen die we zullen testen (paragraaf 4.2.2), stellen we de resultaten van ons onderzoek voor (paragraaf 4.2.3). Conform het aantal operationaliseringingen zullen de resultaten in drie luiken gepresenteerd worden: paragraaf 4.2.3.1 is gewijd aan het effect van de morfologische opbouw van het deelwoord, paragraaf 4.2.3.2 zoomt in op de impact van de afstand van het deelwoordaccent tot het begin van de eindgroep en paragraaf 4.2.3.3 kijkt tot slot naar de impact van de afstand van het deelwoordaccent tot het einde van de eindgroep.

4.2.1 *Status quaestionis*⁷

De accentverdeling in de werkwoordelijke eindgroep heeft zich van in het begin van de wetenschappelijke activiteit rond de woordvolgordevariatie in de werkwoordelijke eindgroep toegespitst op het onderscheid tussen scheidbaar samengestelde en niet-scheidbaar samengestelde deelwoorden (*opgewacht* versus *gewacht*). Scheidbaar samengestelde deelwoorden hebben in tegenstelling tot niet-scheidbaar samengestelde deelwoorden de hoofdklemtoon op het eerste, scheidbare deel (ANS 1997: 606-608) en dat zorgt er blijkens vele onderzoeken voor dat een opvallend verschil in plaatsingsvoorkeur tussen scheidbaar samengestelde en niet-scheidbaar samengestelde werkwoorden vastgesteld kan worden. Zo merken Meeussen & Vanacker (1951: 37) op basis van een aantal romanteksten op dat bij een “deelwoord met preverbium [...] de hoofdvorm [het hulpwerkwoord, GDS] alleen dan achteraan [staat], als er slechts toonloze woorden voorafgaan. [...] Bij gewoon deelwoord integendeel staat de hoofdvorm meest achteraan [...]” Anders geformuleerd: de groene volgorde [PART+AUX] is het frequentst als er sprake is van een niet-scheidbaar samengesteld deelwoord, de rode volgorde [AUX+PART] is het frequentst bij scheidbaar samengestelde deelwoorden, behalve als er ‘toonloze’ woorden aan de eindgroep voorafgaan. Meeussen & Vanacker baseren hun beweringen dan wel op reëel literair taalgebruik, toch is het hen in eerste instantie niet te doen om het verstrekken van kwantitatieve gegevens en statistische verbanden, maar wel om een beschrijving van de verschillende mogelijkheden. De tendens die zij hebben vastgesteld is in later, meer empirisch-kwantitatief onderbouwd onderzoek, verschillende keren bevestigd en herbevestigd (De Schutter 1964 en 1976, De Cubber 1986 en Haeseryn 1990).

Het eerste onderzoek dat De Schutter (1964: 69-70) uitvoert, gebaseerd op een tiental romanteksten uit België en Nederland, resulteert voor clusters die een scheidbaar samengesteld deelwoord bevatten in 74.28% rode volgordes [AUX+PART] (n = 552). Voor clusters zonder scheidbaar samengesteld werkwoord telt De Schutter slechts 33.37% rode volgordes [AUX+PART] (n = 1732). Dat verschil in plaatsingsvoorkeur is erg significant ($\chi^2 = 285.33$, $df = 1$, $p < .0001$). De kans op een rode volgorde [AUX+PART], in vergelijking met de kans op een groene volgorde [PART+AUX] is 5.76 keer groter in een verbale cluster met een scheidbaar samengesteld deelwoord in vergelijking met een cluster zonder scheidbaar

samengesteld deelwoord (o.r. = 5.76, B.I. = 4.63-7.20). Ook De Schutter interpreteert die resultaten in termen van accentuering, maar hij legt wel andere accenten dan Meeussen & Vanacker. Zo herleidt hij het onderscheid tussen de twee types deelwoorden tot het feit dat er na het hoofdaccent op het niet-verbale deel (i.e. het scheidbare deel) van het deelwoord nog “tenminste twee, in zeer veel gevallen drie of zelfs meer niet- of zwakbetoonde lettergrepen na elkaar [komen]: het ideale einde van de bevestigende zin in het Nederlands” (De Schutter 1964: 70). Haeseryn (1990: 53) heeft er o.i. terecht op gewezen dat met een dergelijke formulering niet in te zien is waarom de groene volgorde [PART+AUX], met het een- of tweelettergrepige onbeklemtoonde hulpwerkwoord aan het eind van de eindgroep in dat geval niet minstens even ideaal zou zijn. In later werk zal De Schutter (1996: 215) die formulering wat bijstellen door te zeggen dat twee beklemtoonde syllaben idealiter door twee of drie—niet meer of niet minder—onbeklemtoonde syllaben gescheiden moeten worden.

In het kader van zijn globale beschrijving van de Nederlandse zin doet De Schutter (1976) de analyse uit 1964 nog eens over, nu gebaseerd op een wat meer gediversifieerde dataset, maar nog steeds met een sterk bellettristische inslag. Van de 865 observaties met een dubbele werkwoordsgroep tekent hij 47.41% rode volgordes [AUX+PART] op bij clusters met een scheidbaar samengesteld werkwoord (De Schutters (1976: 233) groep ‘bijw.’) en het hulpwerkwoord *zijn*, *hebben* of *worden* (De Schutters groepen 1 en 2) en slechts 15.67% rode volgordes [AUX+PART] bij clusters zonder scheidbaar samengesteld deelwoord. Dat verschil is ook hier weer statistisch significant ($\chi^2 = 81.14$, $df = 1$, $p < .0001$) en laat zien dat de odds in het voordeel zijn van de rode volgorde [AUX+PART] als het hoofdwkwoord een scheidbaar samengesteld werkwoord is (o.r. = 4.29, B.I. = 3.05-6.07). De verklaring voor het verschil zoekt De Schutter deze keer in het feit dat hulpwerkwoorden met een weinig geprononceerde betekenisstructuur (*hebben*, *zijn*, *worden*) praktisch nooit zelf het groepsaccent dragen, maar wel het liefst zo dicht mogelijk bij het groepsaccent staan, voor zover er voldoende keuzemogelijkheid is. Dat betekent in concreto dat het hulpwerkwoord het liefst voor een scheidbaar samengesteld deelwoord, met het groepsaccent op het scheidbare gedeelte, geplaatst wordt; in de andere gevallen is de keuze principieel onbeperkt. Haeseryn (1990: 53) merkt hier op dat dat problemen

oplevert bij meerledige eindgroepen, “al was het maar omdat er nu altijd ten minste twee hulpwerkwoorden in het geding zijn”.

Belangrijker is echter dat De Schutter (1976) de bespreking van de specifieke invloed die een scheidbaar samengesteld deelwoord uitoefent, openbreekt naar het effect dat een preverbale constituent kan hebben: preverbale constituenten zijn weliswaar geen onderdeel van een complex deelwoord, maar ze kunnen wel, net als het scheidbare deel van het scheidbaar samengestelde deelwoord, een accent dragen. Op die manier betreft De Schutter (1976: 191) primaire bepalingen of directe (inherente) bepalingen in de analyse, zoals adjectieven, substantieven en voorzetselgroepen, die gekenmerkt worden door een nauwe semantische band met het werkwoord (voorbeeldzin (8) is gebaseerd op Haeseryn 1990: 49).

(8) [...] dat zij op reis geweest is.

Dat dergelijke primaire bepalingen, die volgens de vigerende spellingconventies⁸ niet (allemaal) aan het deelwoord vastgeschreven worden, erg vergelijkbaar zijn met de klasse die we traditioneel *scheidbaar samengesteld deelwoord* noemen, blijkt niet alleen uit de accentuering en de semantische verbondenheid met het deelwoord, maar ook uit het feit dat in beide gevallen het niet-verbale deel van het deelwoord gescheiden kan worden (zie verder: Booij & Van Santen 1998: 163-165).⁹ Uit dat alles blijkt dat de grens tussen wat traditioneel scheidbaar deel van een scheidbaar samengesteld werkwoord genoemd wordt en de zgn. (andere) primaire bepalingen die De Schutter op het oog heeft, in feite flinterdun is (althans voor sommige types; zie ANS 1997: 609-610). Toch heeft het orthografische onderscheid aan de grondslag gelegen van heel wat van de eerdere analyses. Het spreekt echter voor zich dat, indien het effect van scheidbaar samengestelde deelwoorden inderdaad begrepen moet worden tegen de achtergrond van een evenwichtige accentverdeling, het ontoelaatbaar is om de onmiddellijke linkercontext van het deelwoord, onafhankelijk van de schrijfwijze, te negeren. In de rest van dit hoofdstuk zal de traditionele grens tussen die beide types van linkercontext geleidelijk aan opgeheven worden. In deze en de twee volgende paragrafen houden we ons echter voorlopig nog aan het traditionele onderscheid.

De twee meest recente onderzoeken die we tot slot nog willen vermelden, zijn die van De Cubber (1986) en Haeseryn (1990). De eerste is gebaseerd op een oorlogsdagboek van een inwoner van Lokeren, geschreven in een regionaal gekleurd Nederlands, waaruit om en bij de 520 relevante attestaties geëxtraheerd werden (het totale aantal verschilt bij De Cubber lichtjes van tabel tot tabel). De verbale clusters met een scheidbaar samengesteld deelwoord vertoonden ook hier weer een sterke neiging voor de rode volgorde [AUX+PART] (77.67%), terwijl de niet-scheidbaar samengestelde deelwoorden bij voorkeur voor het hulpwerkwoord geplaatst werden (64.06%) ($\chi^2 = 51.56$, $df = 1$, $p < .0001$; o.r. = 4.83, B.I. = 3.00-7.81). Het werk van Haeseryn verschilt van de voorgaande onderzoeken, doordat de resultaten niet gebaseerd zijn op corpusmateriaal, maar op experimenteel verkregen materiaal (zie paragraaf 3.2.1 en Haeseryn 1990: 91-133). Omdat wij enkel geïnteresseerd zijn in werkwoordsgroepen met een deelwoord en het hulpwerkwoord *zijn*, *hebben* of *worden* in een bijzin, betrekken we enkel de resultaten voor Haeseryns testitems (B203) en (B106) in de analyse (items uit respectievelijk de puzzeltest en de *maak vragend*-test). Daaruit blijkt dat er op basis van de 500 in experimentele omstandigheden geproduceerde zinnen 55.17% groene eindgroepen [PART+AUX] met een scheidbaar samengesteld werkwoord zijn en 75.73% groene eindgroepen [PART+AUX] zonder scheidbaar samengesteld deelwoord. Dat verschil is opnieuw significant ($\chi^2 = 23.18$, $df = 1$, $p < .0001$; o.r. = 2.53, B.I. = 1.7-3.8).

4.2.2 Hypotheses

Uit het bovenstaande literatuuroverzicht is duidelijk geworden dat de impact van de accentverdeling in de werkwoordelijke eindgroep totnogtoe voornamelijk geoperationaliseerd is in termen van morfologische bouw van het deelwoord, i.e. scheidbaar samengesteld versus niet-scheidbaar samengesteld. Uit de resultaten van die analyses is daarbij eenduidig gebleken dat scheidbaar samengestelde deelwoorden vaker in de rode volgorde [AUX+PART] voorkomen dan niet-scheidbaar samengestelde deelwoorden. Het voorkeursverschil was daarbij in alle onderzoeken zo groot dat telkens met meer dan 99.99% zekerheid besloten kon worden dat het verschil geen toeval is. Het hoeft dan ook niet te verwonderen dat we in deze sectie de volgende hypothese zullen testen:

[HYP 4.1] Scheidbaar samengestelde deelwoorden komen vaker in de rode volgorde [AUX+PART] voor dan niet-scheidbaar samengestelde deelwoorden (paragraaf 4.2.3.1).

Naast de operationalisering in termen van morfologische structuren, zullen we in deze sectie de impact van de accentverdeling in de werkwoordelijke eindgroep ook op een meer fijnkorrelige, specifiek prosodische manier operationaliseren, meer bepaald door te kijken naar de positie van de beklemtoonde syllabes in de werkwoordelijke eindgroep. De hypotheses die aan de grondslag zullen liggen van die specifiekere tests, zijn gebaseerd op het accentueringsprincipe dat in paragraaf 4.1 voorgesteld werd. We brengen voor de duidelijkheid even in herinnering dat dat accentueringsprincipe opgebouwd is rond (i) de evenwichtige verdeling van zinsaccenten in de zin (i.e. elk zinsaccent wordt idealiter voorafgegaan en gevolgd door 2 à 3 onbeklemtoonde syllabes) en (ii) het vrijwaren van het platte-hoedpatroon (i.e. een prosodische grens wordt voorafgegaan door minstens 1 onbeklemtoonde syllabe). Indien we dan voorlopig even aannemen dat er zich net voor de werkwoordelijke eindgroep een zinsaccent bevindt en net na de werkwoordelijke eindgroep ofwel een zinsaccent ofwel een prosodische grens gelokaliseerd is, dan verwachten we het volgende:

[HYP 4.2] Naarmate de afstand (in onbeklemtoonde syllabes) tussen het begin van de werkwoordelijke eindgroep en het deelwoordaccent kleiner wordt, stijgt de voorkeur voor de rode volgorde [AUX+PART] (paragraaf 4.2.3.2);

[HYP 4.3] Naarmate de afstand (in onbeklemtoonde syllabes) tussen het deelwoordaccent en het einde van de werkwoordelijke eindgroep kleiner wordt, stijgt de voorkeur voor de groene volgorde [PART+AUX] (paragraaf 4.2.3.3).

De redenering achter [HYP 4.2] is dat de rode volgorde [AUX+PART], met het onbeklemtoonde hulpwerkwoord vooraan, een accentenbotsing tussen het zinsaccent op het deelwoord (dat zich dus op het eerste, niet-verbale deel van het deelwoord bevindt) en een preverbaal zinsaccent kan vermijden. Vergelijk de minder ‘ideale’ situatie in voorbeeldzin (9b) met die in (9a) (de accentdragende syllabes hebben we onderstreept):¹⁰

- (9a) [...] dat de vereiste inspanningen teveel worden afgewenteld op loontrekkenden en gepensioneerden. (DS)
- (9b) [...] dat de vereiste inspanningen teveel afgewenteld worden op loontrekkenden en gepensioneerden.

De redenering achter [HYP 4.3] is dat de groene volgorde [PART+AUX], met het onbeklemtoonde hulpwerkwoord achteraan, ervoor zorgt dat er voldoende prosodische ruimte gecreëerd wordt om een toonhoogtedaling (die het platte-hoedpatroon afsluit) te realiseren. Vergelijk de minder ‘ideale’ situatie in (10a) met die in (10b) (de accentdragende syllabe hebben we onderstreept):

- (10a) [...] dat ze in Bhutan zwaar werden gediscrimineerd. (DS)
- (10b) [...] dat ze in Bhutan zwaar gediscrimineerd werden.

4.2.3 Resultaten en discussie

4.2.3.1 *Scheidbaar samengesteld versus niet-scheidbaar samengesteld deelwoord*

We beginnen onze onderzoeken naar de impact van de accentverdeling binnen de werkwoordelijke eindgroep met [HYP 4.1]:

- [HYP 4.1] Scheidbaar samengestelde deelwoorden komen vaker in de rode volgorde [AUX+PART] voor dan niet-scheidbaar samengestelde deelwoorden.

Om te verifiëren of scheidbaar samengestelde deelwoorden ook in onze dataset de kans op een rode volgorde [AUX+PART] doen stijgen, hebben we ons materiaal opgesplitst in een groep met werkwoordelijke eindgroepen die een scheidbaar samengesteld deelwoord bevatten en een groep met werkwoordelijke eindgroepen zonder scheidbaar samengesteld deelwoord. De detectie leverde geen problemen op; bij twijfel werd een van de tests in de ANS (1997: 606-608) gebruikt. Daarbij moet wel opgemerkt worden dat er bij sommige complexe deelwoorden twijfel kan bestaan bij de redacteurs van *De Standaard* over het al dan niet aaneenschrijven van de twee delen:

- (11a) [...] in het vaarwater van politieke partijen is terechtgekomen? (DS)

(11b) [...] in het vaarwater van politieke partijen terecht is gekomen?

Uit het voorbeeldenpaar blijkt dat sommige scheidbaar samengestelde deelwoorden niet aan elkaar geschreven hoeven te worden. Dat orthografische verschil houdt geen verband met een of ander prosodisch verschil, zodat de beide zinnen vanuit het onderzoeksperspectief dat in dit hoofdstuk gehanteerd wordt niet van elkaar verschillen. Door de automatische identificatie van de twee types deelwoorden komt het (a)-voorbeeld echter in de groep ‘scheidbaar samengesteld’ terecht en het (b)-voorbeeld in de groep ‘niet-scheidbaar samengesteld’. Met het vooruitzicht dat verderop in dit hoofdstuk de grenzen tussen wat aan elkaar geschreven is en wat niet aan elkaar geschreven is, opgeblazen wordt, hebben we geen pogingen ondernomen om types als het deelwoord in het (b)-voorbeeld op te sporen.

Tabel 4.3 toont de verdeling van de twee types deelwoorden over de responsvariabele:

Tabel 4.3: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type deelwoord.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Scheidbaar samengesteld deelwoord	16.44% (96/584)	83.56% (488/584)
Niet-scheidbaar samengesteld deelwoord	38.37% (693/1806)	61.63% (1113/1806)

Het verschil in verdeling van scheidbaar samengestelde werkwoorden en niet-scheidbaar samengestelde werkwoorden over de groene [PART+AUX] en rode woordvolgorde [AUX+PART] is geheel volgens de verwachting: scheidbaar samengestelde deelwoorden komen opvallend vaker voor in de rode volgorde [AUX+PART] dan niet-scheidbaar samengestelde deelwoorden. Het verschil in distributie is statistisch significant ($\chi^2 = 96.00$, $df = 1$, $p < .0001$), wat betekent dat we voor meer dan 99% zeker kunnen zijn dat er een echt verband is tussen de keuze van de woordvolgorde en de morfologische structuur van het deelwoord. Belangrijker nog dan een statistisch significant verband is de sterkte en de richting van het verband,

uitgedrukt in termen van o.r.'s. De o.r. is in dit geval 0.32 (B.I. = 0.25-0.40), wat betekent dat de waarschijnlijkheid van een groene volgorde [PART+AUX] (versus een rode volgorde [AUX+PART]) zowat drie keer kleiner is als de werkwoordelijke eindgroep een scheidbaar samengesteld deelwoord bevat, vergeleken met een eindgroep zonder scheidbaar samengesteld deelwoord.

Uit dat alles kan men nu concluderen dat de resultaten van eerder onderzoek bevestigd kunnen worden (bevestiging van [HYP 4.1]) en dat het accentueringsprincipe als verklaringsmechanisme alvast in een van zijn aspecten bevestigd lijkt te worden: omdat het accent op het scheidbare, niet-verbale deel van het deelwoord valt, hebben taalgebruikers de neiging om het hulpwerkwoord voorop te plaatsen en zo een accentenbotsing te vermijden met het preverbale materiaal. Het spreekt voor zich dat daarbij aangenomen moet worden dat er voor elke werkwoordelijke eindgroep een zwaar beklemtoond woord staat—meer nog: een woord dat een zinsaccent draagt—en dat de afstand tussen de twee beklemtoonde syllabes kleiner is dan twee syllabes, zodat de neiging om die twee accenten van elkaar te scheiden door het invoegen van het niet- of zwakbetoonde hulpwerkwoord gerechtvaardigd kan worden. Hieruit blijkt nogmaals dat een verificatie van het accentueringsprincipe niet beperkt kan worden tot een analyse van het accentpatroon binnen de werkwoordelijke eindgroep, zoals in vorig onderzoek massaal gebeurd is, maar dat het accentpatroon van de onmiddellijke linker- én rechtercontext in een globale analyse erbij betrokken moet worden. In het vervolg van dit hoofdstuk, wanneer de linker- én rechtercontext in de analyses betrokken worden, zal dan ook moeten blijken of de interpretatie in termen van accentuering gerechtvaardigd is, dan wel of er een andere interpretatie gezocht moet worden.

Vooraleer we daaraan beginnen, willen we in de volgende paragraaf eerst nog nagaan of de groep der scheidbaar samengestelde deelwoorden wel zo homogeen is als totnogtoe aangenomen werd. Bij nadere beschouwing blijkt namelijk dat sommige scheidbare delen uit meer dan 1 syllabe bestaan en, belangrijker, dat sommige van die plurisyllabische scheidbare delen het accent niet op de eerste syllabe hebben. Bekijk de volgende voorbeelden (de syllabe die het deelwoordaccent draagt, is onderstreept):

- (12) [...] dat de verhuur van appartementen en villa's aan de kust in de eerste helft van het jaar is teruggelopen met 10 procent. (DS)

- (13) [...] dat overeengekomen is dat voor 100 miljard begrotingssaneringen gezocht worden. (DS)

In het kader van een principe dat expliciet uitspraken doet over ideale en minder ideale accentpatronen is dat geen onbelangrijk onderscheid: een extra onbeklemtoonde syllabe aan het begin van het deelwoord kan de voorkeur voor de rode volgorde [AUX+PART] nl. drastisch doen dalen (omdat de kans op een accentenbotsing daalt). In de volgende twee paragrafen wordt de impact van de accentverdeling in de werkwoordelijke eindgroep daarom op een specifiek prosodische manier bestudeerd. Paragraaf 4.2.3.2 focust op de afstand tussen het begin van de eindgroep en het deelwoordaccent, paragraaf 4.2.3.3 focust op de afstand tussen het deelwoordaccent en het einde van de werkwoordelijke eindgroep.

4.2.3.2 *Afstand van het deelwoordaccent tot het begin van de eindgroep*

In een tweede stap analyseren we de impact van de accentverdeling in de werkwoordelijke eindgroep niet meer door de deelwoorden categorisch op te splitsen in termen van morfologische structuren (scheidbaar samengesteld versus niet-scheidbaar samengesteld), maar in termen van het aantal onbeklemtoonde syllabes tussen het begin van de eindgroep en het deelwoordaccent. De hypothese is de volgende:

- [HYP 4.2] Naarmate de afstand tussen het begin van de werkwoordelijke eindgroep en het deelwoordaccent kleiner wordt, stijgt de voorkeur voor de rode volgorde [AUX+PART].

Voor de identificatie van het deelwoordaccent werd de prosodische informatie in EGVD 1.4 (2004) geraadpleegd. Bij de berekening van de afstand tussen het deelwoordaccent en het begin van de werkwoordelijke eindgroep werd om voor de hand liggende redenen het aantal syllabes van het hulpwerkwoord, die in de rode volgorde [AUX+PART] kunnen interveniëren tussen het deelwoord en het begin van de eindgroep, niet meegeteld, om een incorporatie van de responsvariabele in de verklarende variabele te vermijden. Indien we dat niet zouden doen, zouden de

intervallen in observaties met een rode volgorde [AUX+PART] steevast een of meer extra onbeklemtoonde lettergrepen bevatten.

Tabel 4.4: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per afstand tussen het begin van de werkwoordelijke eindgroep en het deelwoordaccent.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Afstand = 0 syllabes	15.80% (88/557)	84.20% (469/557)
Afstand = 1 syllabe	37.24% (607/1630)	62.76% (1023/1630)
Afstand = 2 syllabes	47.47% (47/99)	52.53% (52/99)
Afstand = 3 syllabes	43.66% (31/71)	56.44% (40/71)
Afstand = 4 of 5 syllabes	48.48% (16/33)	51.52% (17/33)

Uit tabel 4.4 komt een lineaire, zij het een niet-perfect lineaire trend naar voren: naarmate de afstand tussen het begin van de eindgroep en het deelwoordaccent groter wordt, wordt de rode volgorde [AUX+PART] minder geprefereerd. De trend is statistisch significant ($MH-\chi^2 = 73.44$, $df = 1$, $p < .0001$) en levert een γ -waarde van -0.43 (ASE = 0.04) op. Aangenomen dat er zich in de preverbale linkercontext een zwaar beklemtoond element bevindt, passen die bevindingen perfect in de accentueringshypothese: hoe dichter twee zwaar beklemtoonde syllabes in elkaars buurt liggen, hoe groter de neiging om het hulpwerkwoord tussen die twee syllabes te plaatsen (de rode volgorde [AUX+PART]; bevestiging van [HYP 4.2]), zodat een accentenbotsing vermeden kan worden.

Minder voor de hand liggend is het feit dat de categorie waarbij de afstand 3, 4 of 5 syllabes bedraagt, geen grotere voorkeur voor de groene volgorde [PART+AUX] oplevert, gegeven de eis dat het interval met onbeklemtoonde lettergrepen niet te groot mag zijn. Indien in die gevallen een secundair beklemtoonde syllabe voorafgaat, kan

het accentueringsprincipe behouden blijven, anders dringt een re-evaluatie van dat aspect van het accentueringsprincipe zich op.

De vraag die we ons tot slot nog moeten stellen, is in hoeverre de specifieke operationalisering in termen van beklemtoonde en onbeklemtoonde syllabes (deze paragraaf) verschilt van de operationalisering in termen van morfologische structuren (vorige paragraaf). De resultaten van die beide operationaliseringën werden immers allebei geïnterpreteerd in termen van ‘vermijden van een accentenbotsing aan het begin van de werkwoordelijk eindgroep’. Indien de beide operationaliseringën inderdaad hetzelfde effect van de accentverdeling in de werkwoordelijke eindgroep vatten, dan verwachten we dat de resultaten van die beide analyses (zo veel mogelijk) overlappen. In hoofdstuk 3 hebben we een test voorgesteld die de mate van overlapping van twee factoren statistisch kan bepalen: de kappatetest (κ). De resultaten van de kappatetest laten nu inderdaad zien dat de mate van overeenkomst tussen de operationalisering in morfologische structuren en de specifiek prosodische operationalisering bijna volledig overlappen (de categorieën met 1, 2, 3, 4 of 5 tussenliggende onbeklemtoonde syllabes zijn voor deze analyse samengenomen): $\kappa = 0.96$ (B.I. = 0.95-0.97). Aangezien κ een waarde kan aannemen die kleiner is dan of gelijk is aan 1, waarbij 1 staat voor een perfecte overeenkomst, kunnen we op basis van een kappawaarde van 0.96 dan ook besluiten dat de overeenkomst tussen de twee factoren heel erg groot is, dat de operationalisering van de factor *accentverdeling in de werkwoordelijke eindgroep* in termen van morfologische structuren adequaat is (anders geformuleerd: de groep der scheidbaar samengestelde deelwoorden zijn vanuit prosodisch opzicht voldoende homogeen om als geldige operationalisering van een prosodisch principe te functioneren).

4.2.3.3 *Afstand van het deelwoordaccent tot het einde van de eindgroep*

In de twee vorige paragrafen hebben we ons uitsluitend geconcentreerd op een accentenbotsing vooraan, maar uiteraard kan er zich ook een accentenbotsing voordoen aan het eind van de werkwoordelijke eindgroep. Toen we in tabel 4.4 constateerden dat het vreemd is dat de deelwoorden met het accent op de 4^{de}, 5^{de} of 6^{de} syllabe geen grotere voorkeur vertoonden voor de groene volgorde [PART+AUX], dan is dat niet alleen vreemd omdat dat het interval met onbeklemtoonde lettergrepen

vooraan zou beperken, maar ook omdat die volgorde achteraan voldoende ruimte levert om een prosodische grens te markeren of een accentenbotsing met een zwaar beklemtoonde syllabe in de laatste zinsplaats vermijdt. In tabel 4.5 wordt daarom het aantal onbeklemtoonde syllabes tussen het deelwoordaccent en het einde van de werkwoordelijke eindgroep uitgezet tegenover de woordvolgorde in de werkwoordelijke eindgroep. De hypothese is de volgende:

[HYP 4.3] Naarmate de afstand tussen het deelwoordaccent en het einde van de werkwoordelijke eindgroep kleiner wordt, neemt de voorkeur voor de groene volgorde [PART+AUX] toe.

Bij de berekening van de afstand tussen het deelwoordaccent en het einde van de werkwoordelijke eindgroep werd, om dezelfde reden als in paragraaf 4.2.3.2, het aantal syllabes van het hulpwerkwoord niet meegeteld.

Tabel 4.5: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per afstand van het deelwoordaccent tot het einde van de werkwoordelijke eindgroep.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Afstand = 0 syllabes	39.09% (437/1118)	60.91% (681/1118)
Afstand = 1 syllabe	37.11% (252/679)	62.89% (427/679)
Afstand = 2 syllabes	16.51% (52/315)	83.49% (263/315)
Afstand = 3 syllabes	17.27% (48/278)	82.73% (230/278)

Tabel 4.5 laat een positief lineaire trend zien tussen de proportie rode volgordes [AUX+PART] en de afstand tussen het deelwoordaccent en het einde van de werkwoordelijke eindgroep. De trend is statistisch significant ($MH-\chi^2 = 77.21$, $df = 1$, $p < .0001$) en levert een γ -waarde van 0.29 (ASE = 0.03) op. Het resultaat past opnieuw in het accentueringsprincipe, aangezien bij een korte afstand de groene volgorde [PART+AUX] vaker gekozen wordt (vermijden van een accentenbotsing

achteraan) dan bij een langere afstand (vermijden van een te lang interval met onbeklemtoonde lettergrepen) (bevestiging van [HYP 4.3]). Toch kan men zich hier opnieuw afvragen hoe het komt dat bij een korte afstand de groene volgorde [PART+AUX] niet in een hevigere mate verkozen wordt boven de rode volgorde [AUX+PART], die al bij al de frequentste volgorde blijft. Dat doet opnieuw het vermoeden ontstaan dat er weliswaar een accentueringsprincipe werkzaam is, maar dat het blijkbaar toch niet als enige determinant van de volgorde in de werkwoordelijke eindgroep opgevat mag worden (zie paragraaf 4.5).

4.3 De impact van de accentverdeling in de linkercontext

In deze sectie gaan we na wat de invloed is van de accentverdeling in de linkercontext van de werkwoordelijke eindgroep op de keuze voor de groene [PART+AUX] of rode [AUX+PART] woordvolgorde. Na een overzicht van de belangrijkste literatuur ter zake (paragraaf 4.3.1) en de hypothesen die aan de grondslag zullen liggen van ons onderzoek (paragraaf 4.3.2), bespreken en evalueren we de resultaten (paragraaf 4.3.3). Conform het aantal operationaliseringën zullen de resultaten in vijf luiken gepresenteerd worden: paragraaf 4.3.3.1 is gewijd aan het effect van de definietheid van de laatste preverbale substantiefgroep, paragraaf 4.3.3.2 zoomt in op de impact van de informationaliteit van het laatste preverbale woord, paragraaf 4.3.3.3 kijkt naar de impact van de inherentie van het laatste preverbale woord, paragraaf 4.3.3.4 onderzoekt de invloed van de lengte van het middenstuk en in paragraaf 4.3.3.5 wordt het effect van de afstand van het deelwoordaccent tot het laatste preverbale woordaccent besproken. Paragraaf 4.3.3.6 is ten slotte gewijd aan een vergelijking van de verschillende operationaliseringën van de factor *accentverdeling in de linkercontext*.

4.3.1 Status quaestionis¹¹

Na het effect van de accentverdeling in de werkwoordelijke eindgroep bestudeerd te hebben, zullen we in deze paragraaf de aandacht verschuiven naar de linkercontext van de werkwoordelijke eindgroep. Ook over de invloed van de linkercontext op de woordvolgorde in de werkwoordelijke eindgroep is heel wat geweten. Zo merken Meeussen & Vanacker (1951: 37) op dat bij een “deelwoord met preverbium [...] de

hoofdvorm [het hulpwerkwoord, GDS] alleen dan achteraan [staat], als er slechts toonloze woorden voorafgaan. [...] Bij gewoon deelwoord integendeel staat de hoofdvorm meest achteraan, ook met een naamwoord als onderwerp; de hoofdvorm staat voorop als er tussen onderwerp en deelwoord een ander zinsdeel komt, met sterkere nadruk dan het deelwoord [...].” Zonder het met die woorden te zeggen, interpreteren Meeussen & Vanacker het verband tussen accentuering en woordvolgorde als een duw- en trekmechanisme dat streeft naar een gelijkmatige verdeling van zware accenten, als wat De Schutter (1996) later een relatief correctiemechanisme zou noemen om hoofdaccenten voldoende ver uit elkaar te houden. Net als in paragraaf 4.2.1 geldt ook hier weer dat de vaststellingen van Meeussen & Vanacker niet door kwantitatieve gegevens, laat staan statistische verificatie, ondersteund worden en dat een aantal aspecten relatief vaag gedefinieerd zijn (wat wordt bijvoorbeeld bedoeld met een sterkere nadruk en toonloze woorden?).

Toch worden hun vaststellingen bevestigd door het gros van de daaropvolgende onderzoeken, waaronder dat van De Schutter (1976: 232-235). De Schutter deelt zijn materiaal op in een groep waarbij de eindgroep voorafgegaan wordt door een primaire of direct (inherente) bepaling en een groep zonder voorafgaande direct (inherente) bepaling. Van dergelijke primaire bepalingen is in paragraaf 4.2.1 (zie voorbeeldzin 8) al gezegd dat De Schutter aanneemt dat die een accent dragen. Vanuit het accentueringsprincipe zou men nu verwachten dat de rode volgorde [AUX+PART] de voorkeur heeft indien een primaire bepaling onmiddellijk voorafgaat aan de werkwoordelijke eindgroep, zodat een accentenbotsing tussen het accent van de primaire bepaling en het accent van het deelwoord vermeden kan worden door er het hulpwerkwoord tussenin te plaatsen. De Schutter stelt in zijn data inderdaad vast dat dat het geval is: de voorkeur voor de rode volgorde [AUX+PART] is opmerkelijk groter (67.89%; 74/109) wanneer de eindgroep voorafgegaan wordt door een primaire bepaling—i.c. enkel adjectieven en voorzetselgroepen—in vergelijking met omstandigheden waar geen primaire bepaling aan de eindgroep voorafgaat (44.97% rode volgordes [AUX+PART]; 389/865) (enkel de hulpwerkwoorden in groep 1 en 2 werden in de analyse betrokken, om de vergelijking met het huidige onderzoek optimaal te houden). Het verschil tussen die twee types is erg significant ($\chi^2 = 20.39$, $df = 1$, $p < .0001$) en levert een o.r. op van 2.82 (B.I. = 1.73-4.64).

Naast de invloed van zwaar geaccentueerde inherente elementen gaat De Schutter ook de invloed na van andere zwaar geaccentueerde elementen, zoals indefiniete substantiefgroepen (De Schutter 1976: 234). Een vergelijking van eindgroepen die voorafgegaan worden door een dergelijke indefiniete substantiefgroep en eindgroepen die niet voorafgegaan worden door een indefiniete substantiefgroep levert het verwachte resultaat op: in het eerste geval werden 61.84% rode volgordes [AUX+PART] geteld (47/76), in het tweede geval 32.50% (181/557), hetgeen opnieuw statistisch erg significant is ($\chi^2 = 20.29$, $df = 1$, $p < .0001$); de o.r. maakt ook hier weer duidelijk dat de waarschijnlijkheid van een rode volgorde [AUX+PART] versus de waarschijnlijkheid van een groene volgorde [PART+AUX] bij een voorafgaande indefiniete substantiefgroep 2.82 keer groter is dan bij een ander type linkercontext (B.I. = 1.73-4.64).

In tegenstelling tot De Schutter vindt De Cubber (1986) in zijn dagboekmateriaal geen effect van de factor, Haeseryn (1990) dan weer wel. Omdat wij enkel geïnteresseerd zijn in werkwoordsgroepen met een deelwoord en het hulpwerkwoord *zijn*, *hebben* of *worden* in een bijzin, betrekken we enkel de resultaten voor Haeseryns testitems (A408), (A305), (B203) en (A404) in de analyse.¹² Daaruit blijkt, conform de verwachtingen, dat eindgroepen met een inherente bepaling of een indefiniete substantiefgroep—informatief belangrijke rhematische subjecten en objecten—in 60.19% de rode volgorde [AUX+PART] aannemen (319/530), terwijl andere eindgroepen slechts 33.80% rode volgordes [AUX+PART] aannemen (170/503). Die verschillende verdeling tussen de rode [AUX+PART] en groene [PART+AUX] volgorde is significant ($\chi^2 = 72.1$, $df = 1$, $p < .0001$; o.r. = 2.96, B.I. = 2.28-3.85).

Het laatste onderzoek dat hier vermeld moet worden, komt van Swerts (1998) die op basis van een aantal descriptieve monologen van 5 sprekers, in experimentele omstandigheden tot stand gekomen en door fonetici geannoteerd, de accentueringshypothese tracht te verifiëren. Hij stelt daarbij vast dat de geproduceerde clusters (die allemaal afgesloten worden door een prosodische grens) vaker in de rode volgorde [AUX+PART] voorkomen, indien alleen het voorgaande woord een accent draagt (en het deelwoord niet). Omgekeerd kwamen eindgroepen vaker in de groene volgorde [PART+AUX] voor, indien er enkel een accent ligt op het deelwoord ($\chi^2 =$

6.25, $df = 1$, $p < .02$). Alhoewel deze resultaten de accentueringshypothese overtuigend bevestigen moet er toch voorzichtig mee omgesprongen worden, aangezien ze gebaseerd zijn op erg lage absolute frequenties. Wat Swerts jammer genoeg niet vermeldt, is of er een verband vastgesteld kan worden tussen accentuering en de niet-prosodisch gedefinieerde categorieën *indefiniete substantiefgroep* en *primaire bepaling*. Een dergelijk verband zou namelijk een verhelderend licht kunnen werpen op de accuratesse van de operationalisering van zwaar geaccentueerde lettergrepen in termen van syntactico-semanticke categorieën, zoals dat in de vorige onderzoeken is gebeurd.

Dat laatste geldt evenzeer voor een factor die hier nog niet vermeld is, maar die traditioneel wel in het kader van het accentpatroon in de linkercontext bestudeerd wordt, met name de lengte van het middenstuk, i.e. het aantal woorden of lettergrepen tussen de eerste pool (i.c. het voegwoord *dat*) en de tweede pool (i.c. de werkwoordelijke eindgroep).¹³ In het onderzoek van De Schutter (1976) duikt een constant negatieve trend op tussen de lengte van het middenstuk en de voorkeur voor de groene volgorde [PART+AUX] (merk op dat De Schutter primaire bepalingen en het afgescheiden deel van een voornaamwoordelijk bijwoord tot de tweede pool rekent), d.w.z. hoe langer het middenstuk, hoe minder groene volgordes [PART+AUX]. “De grote voorliefde voor de volgorde nf-pv [groene volgorde [PART+AUX], GDS] in zinnen zonder 3^e zst. [zinsstuk, GDS] kan samenhangen met de afkeer om de neutrale bijzin te laten eindigen op de enige sterke toonkern die zo’n zin bevat (het ‘hoofdwerkwoord’). Naarmate het aantal woord(groep)en in het 3^e zst. groeit, stijgt dus ook de kans dat er in dat zst. een sterk geaccentueerd tegenwicht voor de niet-finiete vorm gevonden wordt. Daardoor gaat dan de tendens om de twee toonkernen zover mogelijk uiteen te plaatsen, de volgorde pv-nf [rode volgorde [AUX+PART], GDS] bevorderen” (De Schutter 1976: 236; zie ook De Schutter 1964: 70-73, De Schutter 1967: 54, De Cubber 1986: 68-69).

4.3.2 Hypotheses

Uit het bovenstaande literatuuroverzicht is duidelijk geworden dat de impact van de accentverdeling in de linkercontext in vroeger onderzoek op een drietal verschillende manieren geoperationaliseerd is: al dan niet aanwezigheid van een indefiniete

substantiefgroep in de linkercontext, al dan niet aanwezigheid van een direct (inherente) of primaire bepaling en lengte van het middenstuk. In deze sectie zullen we die operationalisering recycleren om na te gaan in hoeverre de resultaten van vorig onderzoek bevestigd kunnen worden. Op basis van de operationalisering in termen van al dan niet aanwezigheid van een indefiniete substantiefgroep zal een nieuwe operationalisering voorgesteld worden, meer bepaald in termen informationaliteit (in de paragraaf 4.3.3.2 gaan we dieper in op de verschillen en gelijkenissen tussen de beide operationalisering):

- [HYP 4.4] Een indefiniete substantiefgroep in de onmiddellijke linkercontext doet de voorkeur voor de rode volgorde [AUX+PART] stijgen (paragraaf 4.3.3.1).
- [HYP 4.5] Hoe hoger de informationaliteit van het woord dat zich links van de eindgroep bevindt, hoe groter de voorkeur voor de rode volgorde [AUX+PART] (paragraaf 4.3.3.2).
- [HYP 4.6] Een direct (inherente) of primaire bepaling in de onmiddellijke linkercontext doet de voorkeur voor de rode volgorde [AUX+PART] stijgen (paragraaf 4.3.3.3).
- [HYP 4.7] Naarmate de lengte van het middenstuk toeneemt, des te meer rode volgordes [AUX+PART] gekozen worden (paragraaf 4.3.3.4).

Naast de indirecte operationalisering van de accentverdeling in de linkercontext, zullen we in deze sectie, net zoals in de vorige sectie (paragraaf 4.2), ook een meer fijnkorrelige, specifiek prosodische analyse uitvoeren, waarbij het aantal onbeklemtoonde syllabes tussen het woordaccent van het laatste preverbale woord en het deelwoordaccent geteld zal worden. De hypothese is de volgende (gebaseerd op het accentueringsprincipe; voor meer uitleg: zie paragraaf 4.1):

- [HYP 4.8] Naarmate de afstand tussen het laatste accent in de linkercontext en het begin van de werkwoordelijke eindgroep kleiner wordt, worden meer rode woordvolgordes [AUX+PART] geprefereerd (paragraaf 4.3.3.5).

De redenering achter [HYP 4.8] is dat de rode volgorde [AUX+PART], met het onbeklemtoonde hulpwerkwoord vooraan, een accentenbotsing tussen het zinsaccent

op het deelwoord en het preverbale zinsaccent kan vermijden (vergelijk voorbeeldzinnen (9a) en (9b) hierboven).

4.3.3 Resultaten en discussie

4.3.3.1 Definietheid van de laatste preverbale substantiefgroep

We starten de analyses naar de invloed van de linkercontext op de woordvolgorde met de operationalisering in termen van indefiniete substantiefgroepen. Uit wetenschappelijk onderzoek (zie 4.3.1) is het volgende gebleken:

[HYP 4.4] Een indefiniete substantiefgroep in de onmiddellijke linkercontext doet de voorkeur voor de rode volgorde [AUX+PART] stijgen.

Voor de analyse van het effect van indefiniete substantiefgroepen hebben we ons beperkt tot die substantiefgroepen waarvan het indefiniete nomen onmiddellijk voor de verbale eindgroep staat, onafhankelijk van het feit of dat nomen het hoofd is van de laatste preverbale constituent of niet.

(14a) dat ze een goed werkstuk heeft afgeleverd. (DS)

(14b) dat die beschuldiging aan het adres van een collega geleid heeft tot het ontslag van Holomisa een tiental dagen geleden. (DS)

In het (a)-voorbeeld is het indefiniete nomen het hoofd van de laatste preverbale constituent, in het (b)-voorbeeld niet. Dat voor de analyse hier afgezien is van het syntactische verschil tussen de twee indefiniete nomina ligt ook voor de hand tegen de achtergrond van het accentueringsprincipe: niet zozeer de syntactische functie binnen een zin bepaalt of er een effect is op de woordvolgorde, maar de geaccentueerdheid, en aangezien hier aangenomen wordt dat een indefiniet nomen een zwaar accent draagt, maakt het niet uit wat de syntactische functie ervan is. Dat we ons bij de analyse ook beperkt hebben tot de woorden die onmiddellijk links staan van de werkwoordelijke eindgroep (verder: linkerwoord) verrast eveneens niet. Men verwacht nl. dat een indefiniet nomen dat niet onmiddellijk links van de werkwoordelijke eindgroep gelokaliseerd is niet per se in een accentenbotsing met een vooropgeplaatst deelwoord terecht komt, omdat de mogelijkheid bestaat dat tussenliggend materiaal voor voldoende onbeklemtoonde lettergrepen zorgt om de accentenbotsing te vermijden.

Vanuit de gedachte dat de sequentie van onbeklemtoonde lettergrepen niet al te groot mag zijn, ligt het zelfs voor de hand om in een dergelijke context te verwachten dat net de groene woordvolgorde [PART+AUX] de voorkeur zal wegdragen.

De gegevens in de dataset werden onderverdeeld in een categorie met indefiniete nomina als linkerwoord, een categorie met definiete nomina als linkerwoord en een categorie met andere woordsoorten als linkerwoord. Definiete nomina werden verondersteld geïdentificeerd te kunnen worden aan de hand van een voorafgaand bepaald lidwoord (*de* en *het*), bezittelijke en aanwijzende voornaamwoorden, collectiverende onbepaalde voornaamwoorden, het hoofdtelwoord *beide*, vooropgeplaatste genitieven; ook eigennamen zijn als definitief gecategoriseerd. Indefiniet zijn die substantiefgroepen die ingeleid worden door het onbepaalde lidwoord *een*, niet-collectiverende onbepaalde voornaamwoorden (ANS 1997: 356-375), hoofdtelwoorden met uitzondering van *beide* (ANS 1997: 429-441, 807), vragende voornaamwoorden en persoonlijke voornaamwoorden met algemene referentie; ten slotte zijn ook substantiefgroepen die weliswaar op het eerste gezicht als definitief gecategoriseerd zouden kunnen worden, maar waarvan duidelijk is dat ze categoriaal of generisch gebruikt zijn (er wordt iets gezegd over het type, niet over het token; ANS 1997: 807-812) als indefiniet gecategoriseerd. Nomina kunnen ook een (synchroon identificeerbare) determinator ontberen en traditioneel worden dergelijke nomina als indefiniet beschouwd. Een voorafgaandelijke gepartitioneerde chi-kwadraatanalyse maakt duidelijk dat de verdeling van de zero-substantiefgroepen en de indefiniete substantiefgroepen niet significant van elkaar verschillen, zodat in het vervolg van deze sectie de beide groepen samengenomen worden.

Tabel 4.6: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per woordsoort van het linkerwoord.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Indefiniet nomen	21.77% (150/689)	78.23% (539/689)
Definiet nomen	30.23% (263/870)	69.77% (607/870)
Andere woordsoorten	45.25% (376/831)	54.75% (455/831)

Als we ons eerst enkel concentreren op de cijfergegevens voor de indefiniete en definiete nomina, dan zien we, geheel in lijn met de verwachtingen, dat het voorkomen van een indefiniet nomen dat voor de werkwoordelijke eindgroep gelokaliseerd is, de keuze voor de rode volgorde [AUX+PART] bevordert. Het verschil in plaatsingsvoorkeur onder druk van een voorafgaand definiet nomen vergeleken met een voorafgaand indefiniet nomen is significant ($\chi^2 = 14.13$, $df = 1$, $p < .0002$) en de o.r. van 0.64 (B.I. = 0.51- 0.81) wijst erop dat de kans op een groene volgorde [PART+AUX] versus de kans op een rode volgorde [AUX+PART] ($1/0.64=$) 1.6 keer kleiner is voor indefiniete nomina dan voor de definiete nomina. De resultaten geven aan dat de indeling die door De Schutter is voorgesteld en door andere onderzoekers is overgenomen, steek houdt en bevestigen meer in het bijzonder wat eerder al vastgesteld was: om de zware accenten op het indefiniete nomen en het deelwoord op een comfortabele afstand uit elkaar te plaatsen, wordt de rode volgorde [AUX+PART] geprefereerd (bevestiging van [HYP 4.4]). Het spreekt voor zich dat er ook hier, zoals in paragraaf 4.2, vanuit gegaan moet worden dat indefiniete substantiefgroepen een homogene groep zijn doordat ze meer dan andere constituenten zwaar geaccentueerd zijn (ze zijn discursief nieuw en vragen daarom extra aandacht) en dat de afstand tussen de beklemtoonde syllabes van het deelwoord en het nomen minder dan 2-3 syllabes bedraagt (in paragraaf 4.3.3.6 komen we hierop terug).

Wanneer we de resultaten voor de andere woordsoorten ook nog in de analyse betrekken, dan kan geconstateerd worden dat het verschil tussen de drie categorieën significant is ($\chi^2 = 98.67$, $df = 2$, $p < .0001$; $\lambda = 0.05$, $ASE = 0.01$), meer nog, er is

zelfs een negatief lineaire trend te bespeuren van indefiniete nomina over definiete nomina naar de groep met de andere woordsoorten ($MH-\chi^2 = 95.94$, $df = 1$, $p < .0001$; $\gamma = -0.34$, $ASE = 0.03$). Die lineaire trend kan ook makkelijk ingepast worden in de accentueringshypothese: als men nl. aanneemt dat indefiniete nomina, die discursief nieuw zijn, sneller een zwaar accent krijgen, definiete nomina, die niet nieuw zijn, maar wel informatieel, minder snel een zwaar accent krijgen en andere woordsoorten nog minder in aanmerking komen om beklemtoond te worden (denk maar aan preposities die niet zozeer informatieel zijn, maar veeleer een syntactische functie in de zin hebben), dan is de daling in voorkeur voor de rode volgorde [AUX+PART] perfect in conformiteit met de accentueringshypothese—hoe kleiner de kans op beklemtoning van de laatste preverbale constituent, hoe kleiner de kans op een rode volgorde [AUX+PART]. In de volgende paragraaf gaan we dieper in op die informationaliteitshypothese.

4.3.3.2 *Informationaliteit van het laatste linkerwoord*

In de vorige paragraaf hebben we kunnen vaststellen dat de graad van informationaliteit een effect lijkt te hebben op de woordvolgorde. Als het inderdaad zo is dat de driedeling in tabel 4.6 opgevat mag worden als een gradatie van meer naar minder informatieel, en als het inderdaad zo is dat de graad van informationaliteit correleert met de graad van accentuering, in die zin dat informatieelere woorden een grotere kans hebben om geaccentueerd te worden dan minder informatiele, dan moeten we ons afvragen of er geen betere indeling mogelijk is dan die in tabel 4.6 om het effect van informationaliteit te testen. Het spreekt nl. voor zich dat sommige woordsoorten die nu samen in de categorie ‘andere woordsoorten’ zitten informatieeler zijn dan andere (bv. adjectieven vs. conjuncties). Om de informationaliteitshypothese op een wat fijnkorreligere manier te verifiëren, hebben we daarom het materiaal in tabel 4.6 verder uitgesplitst naar de woordsoort van het laatste preverbale woord: zo vonden we in onze dataset pronomina, conjuncties, preposities, adverbia / adjectieven¹⁴, nomina / verba¹⁵ en telwoorden. Na een eerste gepartitioneerde chikwadraatanalyse, waarbij op zoek gegaan werd naar die categorieën die niet statistisch van elkaar verschillen, bleek dat de 6 onderscheiden categorieën teruggebracht kunnen worden naar drie, meer bepaald nomina / verba / telwoorden (hoog informatieel) versus adjectieven / adverbia (intermediair

informationeel) versus pronomina / conjuncties / preposities (laag informationeel). We verwachten het volgende:

[HYP 4.5] Hoe hoger de informationaliteit van het woord dat zich links van de eindgroep bevindt, hoe groter de voorkeur voor de rode volgorde [AUX+PART].

Tabel 4.7: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per informationaliteitsgehalte van het linkerwoord.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Hoog informationeel	26.47% (413/1560)	73.53% (1147/1560)
Intermediair informationeel	41.20% (262/636)	58.80% (374/636)
Laag informationeel	58.76% (114/194)	41.24% (80/194)

De resultaten in tabel 4.7 laten zoals verwacht een graduele toename van de groene volgorde [PART+AUX] zien. Als met andere woorden aangenomen mag worden dat er een koppeling is van woordsoort aan informatiegehalte, en van informatiegehalte aan kans op accentuering, dan lijken de gegevens in tabel 4.7 de accentueringshypothese te ondersteunen: hoe kleiner de kans op een accent voor de werkwoordelijke eindgroep, hoe kleiner de kans op de rode volgorde [AUX+PART] (MH- $\chi^2 = 107.23$, $df = 1$, $p < .0001$; $\gamma = -0.39$, $ASE = 0.04$) (bevestiging van [HYP 4.5]). Ook hier geldt weer dat daarbij aangenomen moet worden dat het aantal syllabes tussen de twee geaccentueerde syllabes van het deelwoord en het preverbale woord inderdaad kleiner is dan 2-3. Met andere woorden, hoe plausibel de verklaring van het vastgestelde effect ook moge lijken, het is duidelijk dat de accentueringshypothese niet op basis van pragmatico-syntactische criteria getest moet worden, maar via prosodische criteria: ook al bevestigen de resultaten in deze en vorige paragraaf de resultaten van de vorige onderzoeken en lijkt het er met andere woorden op dat een accentueringsprincipe werkzaam is bij de keuze van woordvolgorde, toch moet een

meer gerichte operationalisering uitmaken of er inderdaad een invloed is van de accenten binnen de linkerconstituent. In paragraaf 4.3.3.5 zullen we een gerichtere analyse uitvoeren, maar eerst zal een andere syntactische operationalisering van de invloed van de accentuering binnen de linkercontext onder de loep genomen worden, met name de invloed van de zogenaamde primaire of directe (inherente) bepaling.

4.3.3.3 *Inherentie van de laatste preverbale constituent*

In deze paragraaf analyseren we het effect van de accentverdeling in de linkercontext op basis van de operationalisering in termen van inherentie van de laatste preverbale constituent. We brengen daarbij in herinnering dat van direct (inherente) of primaire bepalingen, die traditioneel gedefinieerd worden als elementen die een sterke semantische band hebben met het hoofdwerkwoord (zie voorbeeldzin (8)), aangenomen wordt dat ze een accent dragen. Tegen de achtergrond van het accentueringsprincipe verwachten we dan ook het volgende:

[HYP 4.6] Een direct (inherente) of primaire bepaling in de onmiddellijke linkercontext doet de voorkeur voor de rode volgorde [AUX+PART] stijgen.

De redenering achter [HYP 4.6] is dat de rode volgorde [AUX+PART], met het onbeklemtoonde hulpwerkwoord tussen het beklemtoonde deelwoord en de beklemtoonde primaire bepaling, een accentenbotsing kan vermijden.

De moeilijkheid aan de operationalisering van de factor *primaire bepaling* ligt in de definatorische onduidelijkheid. Van de primaire bepaling weten we dat De Schutter (1976) ze onderbrengt in het vierde zinsstuk (= het zinsstuk van de werkwoordelijke eindgroep, de tweede pool). Uit de beschrijving van de elementen die bij De Schutter (1976: 192-198) in het derde zinsstuk (= het middenstuk) terechtkomen, kan afgeleid worden dat De Schutter met de primaire bepaling alvast niet de subjecten, objecten en bijwoordelijke bepalingen op het oog heeft. Vermoedelijk komen De Schutters primaire bepalingen overeen met wat de ANS inherente zinsdelen (1997: 1350-1352) noemt: het naamwoordelijk deel van het naamwoordelijk gezegde, de bepaling van gesteldheid, de noodzakelijke bepaling van

hoedanigheid, de noodzakelijke richtings- of plaatsbepaling, de bepaling van maat en het niet-werkwoordelijk deel van een vaste verbinding.

De discussie die hierna volgt, zal zich beperken tot de eerste en de laatste categorie van primaire bepalingen, met name het naamwoordelijk deel van het naamwoordelijk gezegde (4.3.3.3.1) en de werkwoordelijke vaste verbinding (4.3.3.3.2). Voor de andere categorieën, waarbij uit steekproeven gebleken is dat ze erg infrequent zijn in de dataset, nemen we voorlopig aan dat ze hetzelfde effect hebben op de woordvolgorde als de naamwoordelijke delen en het niet-werkwoordelijke deel van de vaste verbinding.

4.3.3.3.1 *Naamwoordelijk deel van het naamwoordelijk gezegde*

We starten onze analyses met een bespreking van het effect van een voorafgaand naamwoordelijk deel van het naamwoordelijk gezegde op de woordvolgorde in de eindgroep. In tabel 4.8 wordt een overzicht gegeven van de verspreiding van de zinnen met een naamwoordelijk deel en zinnen zonder naamwoordelijk deel van het naamwoordelijk gezegde over de twee volgordevarianten. De volgende koppelwerkwoorden zijn in de analyse betrokken: *gebleken, geschenen, geleken, gedocht, gedunkt, geheten, voorgekomen, gegaan, gekomen, gelopen, geraakt, gestaan, gevallen, gezeten* (de semantische criteria in Smedts & Van Belle (1994: 253) en ANS (1997: 1121-1124) werden gebruikt om het onderscheid te maken tussen het zelfstandige, het auxiliaire en het copulatieve gebruik van die werkwoorden):

Tabel 4.8 Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type primaire bepaling (+/- naamwoordelijk deel van het naamwoordelijk gezegde).

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Naamwoordelijk deel	69.42% (193/278)	30.58% (85/278)
Niet-naamwoordelijk deel	28.22% (596/2112)	71.78% (1516/2112)

Ondanks het feit dat het verschil in woordvolgordevoorkeur tussen de twee types linkerwoorden statistisch significant is ($\chi^2 = 188.60$, $df = 1$, $p < .0001$; o.r. = 5.78, B.I. = 4.40-7.58), doorbreken de resultaten in tabel 4.8 toch het verwachtingspatroon. We verwachtten nl. dat de aanwezigheid van een primaire bepaling in de vorm van een naamwoordelijk deel de keuze voor de rode volgorde [AUX+PART] zou verhogen, maar enkel het omgekeerde blijkt waar te zijn (falsificatie van [HYP 4.6]).

Nu kan men uiteraard opperen dat het accentueringseffect van de primaire bepaling voornamelijk speelt bij een hoog informatiegehalte en dat het ontbrekende effect in tabel 4.8 met andere woorden te wijten zou kunnen zijn aan een mogelijk groot aandeel van laag-informationele naamwoordelijke delen. Uit een gestratificeerde CMH-analyse blijkt echter dat daar niks van aan is: zelfs indien men zou controleren voor graad van informatieel gehalte (zoals in tabel 4.7), dan blijft het globale effect behouden en werden geen redundantie- of interactie-effecten ontdekt.

Dat de resultaten in de bovenstaande tabel afwijken van wat verwacht werd, hoeft echter bij nader inzien niet echt te verbazen: in paragraaf 1.3 is er nl. al op gewezen dat er een verschil in voorkeur lijkt te zijn tussen werkwoordelijke en naamwoordelijke gezegdes, getuige het volgende ANS-citaat: “De plaatsing in de eindgroep van het voltooid en het passief deelwoord als zelfstandig werkwoord is in principe vrij, althans in zuiver werkwoordelijke gezegdes. (Voor naamwoordelijke gezegdes zie men 2.4.6/7 en 18.5.2.3/1, *Opmerking*)”. Uit dat citaat blijkt dat er toch wat speciaals aan de hand lijkt te zijn met de graad van variatiemogelijkheden in de werkwoordelijke eindgroep met een naamwoordelijk gezegde. Nu moet daar onmiddellijk bij gezegd worden dat de ANS-redactie daarbij voornamelijk gedacht heeft aan adjectivisch gebruikte deelwoorden (blijkens de passages waaraan in het citaat gerefereerd wordt), terwijl de naamwoordelijke gezegdes in tabel 4.8 onverdachte deelwoorden bevatten. In hoofdstuk 5 zal hoe dan ook blijken dat er toch een duidelijk preferentieverschil is tussen naamwoordelijke gezegdes als globale categorie en werkwoordelijke gezegdes.

We vatten samen: onder de assumptie dat elk inherent zinsdeel een zwaar accent draagt en vandaar verondersteld wordt de rode volgorde [AUX+PART] uit te lokken, stellen we in onze data vast dat dat alvast niet geldt voor het naamwoordelijk deel van

het naamwoordelijk gezegde als primaire, inherente bepaling. Dat kan o.i. twee zaken betekenen: (i) de hypothese dat een zwaar accent de keuze voor de woordvolgorde kan bepalen, is weerlegd. Dat hoeft uiteraard niet te betekenen dat de accentueringshypothese in haar totaliteit afgewezen moet worden: zo kan bijvoorbeeld het naamwoordelijke deel toch nog als een zwaar beklemtoonde constituent opgevat worden, maar wordt het effect van de accentuering geneutraliseerd door het effect van een ander, krachtiger (semantisch) mechanisme; (ii) de assumptie dat inherente zinsdelen zwaar beklemtoond zijn, moet herbekeken worden (op zijn minst voor het naamwoordelijk deel van het naamwoordelijk gezegde).

4.3.3.3.2 *Niet-werkwoordelijk deel van een werkwoordelijke vaste verbinding*

In een tweede stap willen we, zoals aangekondigd, nagaan of het effect van het niet-werkwoordelijke deel van een werkwoordelijke vaste verbinding de accentueringshypothese wél bevestigt. De zoektocht naar de vaste verbindingen in onze dataset bleek bemoeilijkt te worden doordat de term *vaste verbinding* in standaardwerken en in de vakliteratuur veelal impressionistisch gedefinieerd is. Een gevolg daarvan is dat de term niet overal dezelfde lading dekt—“Je kunt geen artikel over vaste verbindingen lezen of de auteur geeft aan welke definitie hij/zij hanteert, of de auteur heeft zijn/haar eigen variant [...]” (Everaert 1993: 4; zie ook Moon 1998: 2-5)—en dat naar de conceptuele inhoud van de vaste verbinding verwezen wordt met verschillende termen—*idioom, fraseologisme, zegswijze, spreekwoord, uitdrukking*,... (zie bv. Verstraten 1992: 19-24 en Everaert 1993: 4-5).

Door die definatorische en terminologische onduidelijkheid lag de operationalisering van het concept *vaste verbinding* niet voor de hand. Voor de studie van het effect van de vaste verbinding op de woordvolgorde in de eindgroep is er daarom voor geopteerd om de term *vaste verbinding* ruim te definiëren, met name als een overkoepelende term voor gefixeerde patronen van minstens twee woorden (i.c. een werkwoord en een niet-werkwoord). Dat betekent met andere woorden dat we met de verzameling van vaste verbindingen zowel pure idiomen, constructionele idiomen (Booij 2002: 302) als collocaties¹⁶ (Sinclair 1991) op het oog hebben. De opgesomde types vaste verbindingen hebben met elkaar gemeen dat de onderdelen ervan door de taalgemeenschap als een min of meer vaststaande eenheid beschouwd worden, maar

verschillen van elkaar door de graad van semantische compositionaliteit. Het moet van bij aanvang duidelijk zijn dat het er ons in deze studie niet om te doen is de verschillende types vaste verbinding te identificeren, noch om de verzameling van de vaste verbindingen intern te analyseren, maar wel om de verzameling in haar totaliteit op een objectieve en consistente manier af te bakenen, zodat het mogelijk is om voor de eindgroepen in onze dataset te beslissen of ze al dan niet deel uitmaken van een vaste verbinding.

Kwantitatief-statistische detectie van werkwoordelijke vaste verbindingen

Veeleer dan een set van linguïstische criteria op te stellen, zoals in het verleden vaak gebeurd is (zie bv. Verstraten 1992), zullen we vaste verbindingen op een kwantitatief-statistische wijze trachten te identificeren. Als we er namelijk vanuit gaan dat vaste verbindingen gekenmerkt worden door de gefixeerde combinatie van een aantal woorden en als we *gefixeerdheid* gelijkshakelen met *recurrentie*—een sequentie van woorden wordt pas als gefixeerd beschouwd als de sequentie regelmatig terugkeert—dan kunnen we aan de hand van frequentiegegevens trachten uit te maken of een bepaald patroon een vaste verbinding is of niet. Echter, zo kan men redeneren, “aantonen dat een bepaalde woordcombinatie statistisch gezien veel voorkomt, is iets anders dan aantonen dat iets een vaste verbinding is [...]” (Everaert 1993: 20). Everaert illustreert zijn standpunt met het volgende voorbeeldenpaar:

(15a) De minister heeft voorgesteld.

(15b) De vrouw heeft voorgesteld.

Met onze kennis van de wereld lijkt het voor de hand te liggen dat de (a)-voorbeeldzin frequenter zal voorkomen dan de (b)-zin en daarom beweert Everaert dat met een puur statistische aanpak de eerste zin sneller als vaste verbinding beschouwd zal worden dan de tweede (terwijl daar vanuit een puur linguïstisch perspectief geen reden toe is).

Wij volgen Everaert in zijn bedenking dat het vergelijken van absolute frequenties geen goede methode is om vaste verbindingen op het spoor te komen, maar we gaan niet akkoord met de stelling dat een statistische aanpak geen soelaas zou kunnen bieden voor het probleem van de identificatie van vaste verbindingen. De statistiek heeft namelijk meer te bieden dan enkel descriptieve technieken, zoals het

vergelijken van absolute frequenties: aan de hand van inferentieel-statistische technieken, zoals die in dit en het vorige hoofdstuk al ruimschoots aan bod gekomen zijn, kan men nagaan of een bepaalde combinatie van woorden significant vaak samen voorkomt. Dat die significantie niet enkel bepaald kan worden op basis van de frequentie van het patroon, maar dat ook de frequentie waarmee elk van de constituerende delen in andere contexten voorkomt een rol zal moeten spelen, is evident.

Een van de inferentiële technieken die frequent gebruikt wordt om de significantie van samen-voorkomen (i.e. de mate van gefixeerdheid van het patroon of de mate van lexicale attractie tussen de onderdelen van het patroon) te bepalen, is de *log-likelihood ratio test* (LLR), waarvan vergelijkende studies hebben kunnen aantonen dat het een van de betrouwbaarste statistische tests is voor de huidige doelstelling (andere vaak gebruikte lexicale associatiematen zijn *Mutual Information Index* en *Fisher's Exact*; Dunning 1993, Kilgarriff 2001). Om te testen welke sequenties van woorden in onze dataset als vaste verbinding kunnen worden beschouwd, zullen we daarom de LLR-test gebruiken. Meer bepaald zal voor elk patroon dat bestaat uit het woord dat onmiddellijk links van de werkwoordelijke eindgroep (verder: linkerwoord) staat en het deelwoord van de werkwoordelijke eindgroep de LLR-waarde bepaald worden (en op basis daarvan de graad van significantie en gefixeerdheid).

Dat we ons op die twee elementen gaan baseren om uit te maken of we met een vaste verbinding te maken hebben, mag dan al op het eerste gezicht vreemd lijken, het is het allerminst. Laat ons even in herinnering brengen dat we hier het effect van het niet-werkwoordelijke deel van een vaste verbinding, waarvan verondersteld wordt dat het een zwaar accent draagt, op de keuze voor de woordvolgorde willen bestuderen. Om uit te maken of de zin waarin een werkwoordelijke eindgroep gevonden is, een vaste werkwoordelijke verbinding bevat, moeten we in eerste instantie nadenken over de mogelijke posities waar de delen van zo'n vaste verbinding zich in de zin kunnen bevinden. Voor het werkwoordelijke deel van de vaste verbinding is dat alvast duidelijk: het deelwoord dat zich in de werkwoordelijke eindgroep bevindt (andere werkwoorden zijn er niet in de zin, behalve het hulpwerkwoord, waarvan op voorhand al gezegd kan worden dat het niet inherent deel uitmaakt van de verbinding; zie bv. in

O.T.T.). De detectie van het niet-werkwoordelijke deel van de vaste verbinding is minder voor de hand liggend. Toch kunnen de mogelijke posities waar het niet-verbale deel zich kan bevinden op basis van onze kennis over de Nederlandse syntaxis vrij snel beperkt worden tot de positie net voor de werkwoordelijke eindgroep (vandaar dat De Schutter ze ook als lid van het vierde zinsstuk beschouwt, het zinsstuk met o.a. de werkwoordelijke elementen). Het spreekt voor zich dat het niet-verbale deel van een vaste verbinding over het algemeen uit meer dan één woord bestaat en dat daarom de keuze voor het onmiddellijke linkerwoord en niet voor de twee, drie, vier,... laatste preverbale woorden enige verklaring vergt. In eerste instantie moet gezegd worden dat een uitbreiding naar patronen met meer dan twee woorden technisch gezien wel mogelijk is, maar een dermate verhoging van de rekenkracht vergt (tijdens de zoekopdracht én tijdens de statistische verwerking) dat het voorlopig onmogelijk is gebleken om een uitbreiding tot stand te brengen. In tweede instantie, rekening houdend met de beperking tot twee woorden waarvan er een alvast het deelwoord moet zijn, is het zaak om ervoor te zorgen dat de keuze voor een van de woorden van het niet-werkwoordelijke deel van de vaste verbinding optimaal is, i.e. dat het geselecteerde woord een goede, informatieve vertegenwoordiger is van het niet-werkwoordelijke gedeelte, zodat de berekening van de LLR-waarde gebruik kan maken van de woorden die echt typerend zijn voor de verbinding. Functiewoorden zijn dat over het algemeen niet, omdat ze in zoveel verschillende contexten voorkomen en daarom minder kenmerkend zijn voor het niet-verbale deel van een vaste verbinding. Inhoudswaarden zijn daarom veel betere, typische vertegenwoordigers van het niet-werkwoordelijke deel. Als dan ook nog uit een steekproef op een gedeelte van de dataset¹⁷ blijkt dat het laatste woord voor de werkwoordelijke eindgroep dat deel uitmaakt van een vaste verbinding bijna altijd een inhoudswoord is, dan is de beperking van de LLR-analyse tot het onmiddellijke linkerwoord daarmee onmiddellijk gerechtvaardigd.

Om de LLR succesvol te kunnen uitvoeren, hebben we van elk linkerwoord-deelwoord-paar (verder bigram¹⁸ genoemd) in onze dataset de frequentie-informatie nodig. In een eerste stap werden met behulp van een PERL-script de relevante bigrammen in onze dataset geïdentificeerd (daarbij moest onder meer rekening gehouden worden met het feit dat het bigram doorbroken kon worden door een hulpwerkwoord). In een tweede stap werden de bigrammen gelemmatiseerd, i.e. elk

woord in elk van de bigrammen wordt herleid naar de basisvorm (zoals dat gevonden kan worden in een woordenboek; voor werkwoorden is dat bijvoorbeeld de infinitief).¹⁹ Op die manier zijn we in staat om snel en efficiënt op zoek te gaan naar de vaste verbindingen onder de bigrammen, los van het feit of de onderdelen van de verbinding geflecteerd zijn en welk type flexie gebruikt is; we zijn nl. niet enkel geïnteresseerd in de bigrammen zoals die voorkomen in onze dataset, maar in alle verwante bigrammen die op dezelfde basisvorm teruggevoerd kunnen worden. Aangezien het CGN (R5) gelemmatiseerd is (zie hoofdstuk 2 voor details over de structuur en de inhoud van het corpus), hebben we gebruik gemaakt van dat corpus om de frequenties van de bigrammen op te sporen. Daarbij is ook rekening gehouden met het feit dat de positie van het werkwoord en de niet-verbale vertegenwoordiger van het bigram niet altijd respectievelijk rechts en links hoeft te zijn, maar dat bijvoorbeeld in sommige hoofdzinnen het werkwoord links van het niet-verbale deel staat.

Van de 2390 bigrammen in de dataset bleken er op die manier maar liefst 2231 unieke (gelemmatiseerde) bigrammen te zijn (93.35% van het totale aantal bigrammen). Naast de frequentie-informatie van de verschillende bigrammen, zijn ook de individuele frequentiegegevens van de constituerende delen van de bigrammen en de totale corpusfrequentie nodig om de LLR succesvol te kunnen berekenen. De eigenlijke LLR-berekening ziet er als volgt uit:

$$2 ((a (\ln(a))) + ((B-a) (\ln(B-a))) + ((A-a) (\ln(A-a))) + (((C-A)-(B-a)) (\ln((C-A)-(B-a)))) - (B (\ln(B))) - (A (\ln(A))) - (((A-a)+((C-A)-(B-a))) (\ln(((A-a) + ((C-A)-(B-a)))))) - ((C-A) (\ln(C-A))) + (C (\ln(C))))$$

De letters in de formule staan voor de frequenties van het bigram (a), de frequentie van het werkwoordelijke gedeelte van het te onderzoeken patroon (A), de frequentie van het niet-verbale gedeelte van het te onderzoeken patroon (B) en de totale corpusfrequentie (C). Aan de hand van die basisgegevens kan nu afgeleid worden dat de mathematische bewerking:

- (A-a) staat voor ‘de frequentie van het verbale gedeelte van het te onderzoeken patroon (A) zoals dat voorkomt buiten het te onderzoeken patroon (a)’.

- (B-a) staat voor ‘de frequentie van het niet-verbale gedeelte van het te onderzoeken patroon (B) zoals dat voorkomt buiten het te onderzoeken patroon (a)’.
- (C-A) staat voor ‘de frequentie van alle woorden in het corpus (C) behalve het verbale gedeelte van het te onderzoeken patroon (A)’.
- ((C-A)-(B-a)) staat voor ‘de frequentie van alle woorden in het corpus (C) behalve het verbale gedeelte (A) en behalve de frequentie van alle niet-verbale gedeeltes (B) van het te onderzoeken patroon buiten de context van het bigram (a)’.
- ((A-a)+((C-a)-(B-a))) staat voor ‘de frequentie van het verbale gedeelte van het te onderzoeken patroon (A) zoals dat voorkomt buiten het te onderzoeken patroon (a), samen met de frequentie van alle woorden in het corpus (C) behalve het verbale gedeelte (A) en behalve de frequentie van alle niet-verbale gedeeltes (B) van het te onderzoeken patroon buiten de context van het bigram (a)’.

In tabel 4.9 wordt in tabellarische vorm een overzicht gegeven van alle benodigde gegevens, zowel van de basisgegevens als van de afgeleide gegevens; X staat daarbij voor ‘verbaal deel van het te onderzoeken patroon’, Y voor ‘niet-verbaal deel van het te onderzoeken patroon’; \hat{X} en \hat{Y} staan voor de negatie van X en Y:

Tabel 4.9: Basisgegevens voor de berekening van de lexicale attractie met behulp van LLR

	X	\hat{X}	Totaal
Y	a	B-a	B
\hat{Y}	A-a	(C-A)-(B-a)	(A-a)+((C-A)-(B-a))
Totaal	A	C-A	C

Op basis van de frequentiegegevens in de tabel wordt voor elk bigram de LLR-berekening²⁰ uitgevoerd. Dat levert voor elk bigram een score op aan de hand waarvan de statistische significantie bepaald kan worden: hoe hoger de LLR-score, hoe kleiner

de p-waarde (i.e. hoe kleiner de kans dat een dergelijk resultaat te wijten is aan het toeval).

Om een en ander wat aanschouwelijker te maken, geven we de berekening van de mate van lexicale attractie tussen *aandacht* (n = 414) en *besteden* (n = 232); de twee woorden komen 35 keer voor in elkaars onmiddellijke buurt en de totale corpusfrequentie is 4146501:

$$\begin{aligned}
 & 2 ((35 (\ln(35))) + (379 (\ln(379))) + (197 (\ln(197))) + (4145890 (\ln((4145890)))) \\
 & - (414 (\ln(414))) - (232 (\ln(232))) - (4146087 (\ln(4146087))) - (4146269 \\
 & (\ln(4146269))) + (4146501 (\ln(4146501)))) \\
 & = 451.09
 \end{aligned}$$

Het patroon *aandacht besteden* krijgt aldus een LLR-waarde van 451.09. Aangezien van de resultaten van een LLR-test geweten is dat ze een chi-kwadraatverdeling hebben, kan nu gemakkelijk een p-waarde berekend worden. Aangezien de p-waarde kleiner blijkt te zijn dan 0.05 (en zelfs kleiner dan 0.0001), kunnen we zeggen dat het onderzochte patroon statistisch significant is, dat wil zeggen dat de kans dat een dergelijke extreme LLR-waarde te wijten is aan het toeval kleiner is dan 0.0001%. Hieruit kunnen we dan ook besluiten dat *aandacht besteden* als een statistische collocatie beschouwd moet worden.

Nu moet daar onmiddellijk bij opgemerkt worden dat een p-waarde kleiner dan 0.05 weliswaar altijd wijst op statistische significantie, maar in het kader van de huidige test kan dat twee zaken betekenen: ofwel komt het niet-verbale gedeelte significant vaak voor in de buurt van het werkwoord (scenario A) ofwel komt het niet-verbale gedeelte significant niet vaak voor in de buurt van het werkwoord (scenario B). In dat laatste geval komt het niet-verbale gedeelte significant vaker voor bij andere woorden dan bij het targetwerkwoord. Het spreekt voor zich dat wij voor deze studie enkel geïnteresseerd zijn in scenario A en daarom werd de volgende vergelijking gebruikt om voor elke collocatie uit te maken of we met scenario A of met scenario B te doen hadden (de letters staan opnieuw voor dezelfde frequentiegegevens als in tabel 4.10):

$$(a/A) > ((B-a)/(C-A))$$

Het eerste deel van de vergelijking geeft de relatieve proportie weer van het aantal keer dat het niet-verbale gedeelte van het te onderzoeken patroon als onmiddellijke buur van het werkwoord van het te onderzoeken patroon voorkomt, het tweede deel geeft de relatieve proportie weer van het aantal keer dat het niet-verbale gedeelte niet als onmiddellijke buur van het targetwerkwoord gebruikt wordt. Als het linkerdeel groter is dan het rechterdeel, dan is er sprake van een positieve trend (scenario A). Voor *aandacht besteden* bleek dat het geval te zijn ($(35/232) > (379/4146269) = 0.151 > 0.0000914$).

Overzicht van de meest gefixeerde bigrammen

Voor elk patroon dat in de dataset voorkwam, werd op die manier op basis van de CGN-frequenties (R5) een LLR-score berekend (tussen 0 en $+\infty$), waarbij we ervan uitgaan dat die iets zegt over de graad van gefixeertheid van het bigram (in het kader van ons onderzoek naar vaste verbindingen). Van de 2390 bigrammen in de dataset bleek ongeveer een kwart (24.44%; $n = 584$) een statistische collocatie te zijn ($p < .05$). Het spreekt voor zich dat het te veel ruimte zou vergen om alle bigrammen met hun LLR-waarde op te sommen, maar in tabel 4.10 hebben we alvast de top-20 van de meest gefixeerde bigrammen opgesomd:

Tabel 4.10: Top-20 van de meest gefixeerde bigrammen in de dataset.

Niet-verbaal deel	Verbaal deel	LLR-waarde
niet	weten	7287.32
bezig	zijn	1800.46
vraag	stellen	1298.27
goed	zijn	1158.93
door	worden	1124.87
ervan	overtuigen	814.28
rekening	houden	699.59
rol	spelen	683.94
er	zien	606.51
voor	zorgen	566.68
aan	geven	503.56
aandacht	besteden	451.09
op	wijzen	446.78
moeilijk	zijn	436.91
het	gewoon	425.03
ze	zien	418.95
ver	gaan	395.46
bij	betrekken	386.20
beslissing	nemen	368.02
niet	doen	354.24

Uit tabel 4.10 blijkt al onmiddellijk dat op basis van de LLR-test een aantal significante patronen opduiken waarvan we vanuit een traditioneel linguïstisch kader niet onmiddellijk zouden zeggen dat ze vaste verbindingen zijn, zoals de combinaties van werkwoorden met vaste voorzetsels (*zorgen voor*) of een pronomen dat als subject of object tot de argumentstructuur van een werkwoord behoort (*ze zien*). De vraag dringt zich dan ook op of er niet een extra controlemechanisme ontwikkeld moet worden om de resultaten van de statistische test te filteren op linguïstische relevantie. Het probleem met een dergelijk linguïstisch controlemechanisme is dan weer dat niet op voorhand criteria bedacht kunnen worden op basis waarvan een statistische collocatie als een ‘echte’ vaste verbinding mag worden beschouwd—met de statistische aanpak hebben we nu net getracht om onafhankelijk van de linguïstische criteria vaste verbindingen op te sporen. Een mogelijk linguïstisch controlemechanisme zou er wel in kunnen bestaan dat de statistische resultaten

geclassificeerd worden volgens woordsoort. De hypothese achter die opsplitsing is dat werkwoordelijke vaste verbindingen typisch bestaan uit minstens een werkwoord en een nomen; statistische collocaties met een ander niet-verbaal gedeelte vormen minder typische vaste verbindingen. Indien we ervan uitgaan dat die hypothese klopt en de top-20 in tabel 4.10 dientengevolge zouden opsplitsen volgens woordsoort, krijgen we het volgende resultaat:

Tabel 4.11: Top-20 van de meest gefixeerde bigrammen in de dataset, gecontroleerd voor woordsoort.

Gefixeerde bigrammen met een nominaal niet-verbaal deel	Gefixeerde bigrammen met een adjectivisch of adverbiaal niet-verbaal deel	Gefixeerde bigrammen met een prepositioneel niet-verbaal deel
vraag stellen (1298.27)	niet weten (7287.32)	door worden (1124.87)
rekening houden (699.59)	bezig zijn (1800.46)	voor zorgen (566.68)
rol spelen (683.94)	goed zijn (1158.93)	aan geven (503.56)
aandacht besteden (451.09)	ervan overtuigen (814.28)	op wijzen (446.78)
beslissing nemen (368.02)	er zien (606.51)	bij betrekken (386.20)
gebruik maken (316.60)	moeilijk zijn (436.91)	van overtuigen (342.94)
afpraak maken (298.02)	ver gaan (395.46)	voor betalen (282.68)
beslag nemen (273.22)	niet doen (354.24)	op aandringen (209.15)
beschikking stellen (265.51)	erin slagen (319.54)	aan werken (202.38)
geld kosten (259.75)	anders zijn (276.64)	in slagen (191.37)
fout maken (252.61)	nooit zien (262.99)	aan verbinden (186.04)
steek laten (242.89)	niet begrijpen (229.63)	door treffen (65.94)
werk zijn (240.11)	niet zeggen (229.44)	aan overhouden (58.44)
knoop doorhakken (192.91)	goed doen (207.53)	aan schenken (23.51)
antwoord geven (189.89)	binnen komen (203.73)	op nahouden (13.83)
keuze maken (178.82)	anders kunnen (196.53)	uit trekken (11.05)
stand komen (177.30)	niet geloven (187.44)	voor straffen (9.85)
tafel leggen (138.71)	genoeg zijn (172.73)	in betrekken (4.28)
ontslag nemen (134.45)	verantwoordelijk zijn (161.60)	
uitspraak doen (133.55)	ertoe leiden (130.11)	

Gefixeerde bigrammen met een pronominaal niet-verbaal deel	Gefixeerde bigrammen met een voegwoordelijk niet-verbaal deel
het gewoon (425.03)	dat werken (139.84)
ze zien (418.95)	dat luisteren (5.55)
het doen (333.51)	
<u>zich vergissen (133.01)</u>	
dat aankondigen (109.87)	
hij slaan (107.16)	
niets doen (100.28)	
het zeggen (75.09)	
hem helpen (57.52)	
alles verlopen (50.52)	
dat bevestigen (48.92)	
ze gebruiken (46.32)	
hen spreken (45.23)	
hen hebben (40.22)	
iets veranderen (36.88)	
niets opleveren (32.81)	
ze vergeten (32.77)	
<u>zich bewust (31.23)</u>	
die vallen (30.47)	
ze trouwen (30.25)	

In eerste instantie moet gezegd worden dat voor de meeste woordsoortcategorieën in tabel 4.11 geldt dat de top-20 slechts de toplaag is van het totale aantal significante bigrammen. Voor de volledigheid vermelden we hier kort de absolute en relatieve cijfers van het aantal significante bigrammen per woordsoort: voor de nomina is dat 333/1558 (21.37%), voor de adjectieven/adverbia 177/636 (27.83%), voor de preposities 22/26 (84.62%), voor de pronomina 57/162 (35.19%) en voor de conjuncties 2/6 (33.33%). Vooral het hoge percentage significante bigrammen met een prepositie valt op. Wanneer we het lijstje met de preposities bekijken, dan kan geconstateerd worden dat de LLR-test in staat blijkt te zijn om zgn. vaste voorzetsels te detecteren.

Bij een onderlinge vergelijking van de lijsten in tabel 4.11 stellen we verder nog vast dat enkel het lijstje met de nomina ‘echte’ vaste verbindingen lijkt te bevatten

(misschien met uitzondering van *zich bewust in zich ergens bewust van zijn*). We zeggen *lijkt te bevatten*, omdat een dergelijk oordeel uiteraard enkel op basis van onze eigen intuïties uitgesproken kan worden en op basis van wat vermoedelijk traditioneel als prototypische vaste verbinding wordt beschouwd. Om te verifiëren of onze intuïties inderdaad overeenkomen met de traditionele opvattingen, zijn we in twee standaardnaslagwerken, EGVD 1.4 (2004) en Van Dale groot woordenboek van hedendaags Nederlands (GWHN, 1991), nagegaan welke van de statistische collocaties in tabel 4.11 opgenomen zijn als vaste verbinding. Daaruit blijkt dat de bigrammen met adjectieven, preposities, pronomina en voegwoorden inderdaad bijna nooit als vaste verbindingen beschouwd worden. Een aantal twijfelgevallen kan gevonden worden in het verplicht reflexief voornaamwoord en in een aantal adverbiale bepalingen (de onderstreepte bigrammen in tabel 4.11 zijn in EGVD 1.4 (2004) of GWHN (1991) als vaste verbinding gecategoriseerd²¹). Uiteraard betekent dat niet perse dat de statistische collocaties met een niet-nomen als niet-verbaal gedeelte niet als vaste verbinding beschouwd mogen worden. Het enige wat we uit de opzoeking in EGVD 1.4 (2004) en GWHN (1991) kunnen besluiten, is dat de traditionele linguïstische opvatting van wat een vaste verbinding is, overeen lijkt te komen met de vereiste dat het niet-verbale gedeelte minstens een nomen moet bevatten. Toch moet ook voor de lijst met de nomina in tabel 4.11 vastgesteld worden dat niet alle statistische collocaties als vaste verbinding opgenomen zijn in een van de twee referentiewerken (*aandacht besteden*). Omdat op voorhand niet vastgesteld kan worden welke van de voorgestelde operationaliseringen (de puur statistische of de linguïstisch gecontroleerde statistische aanpak) het beste aansluit bij wat een vaste verbinding is en welke operationalisering in het kader van dit onderzoek tot de meest optimale resultaten leidt, zullen we in het vervolg van deze paragraaf de beide benaderingen blijven hanteren en zullen aan het eind de resultaten die de benaderingen opleveren, vergeleken en geëvalueerd worden.

De invloed van een werkwoordelijke vaste verbinding op de keuze voor de groene of de rode volgorde

Laten we vooraleer we verder gaan even alles op een rijtje zetten: we zijn de onderzoekspiste naar een adequate operationalisering van het concept *vaste verbinding*

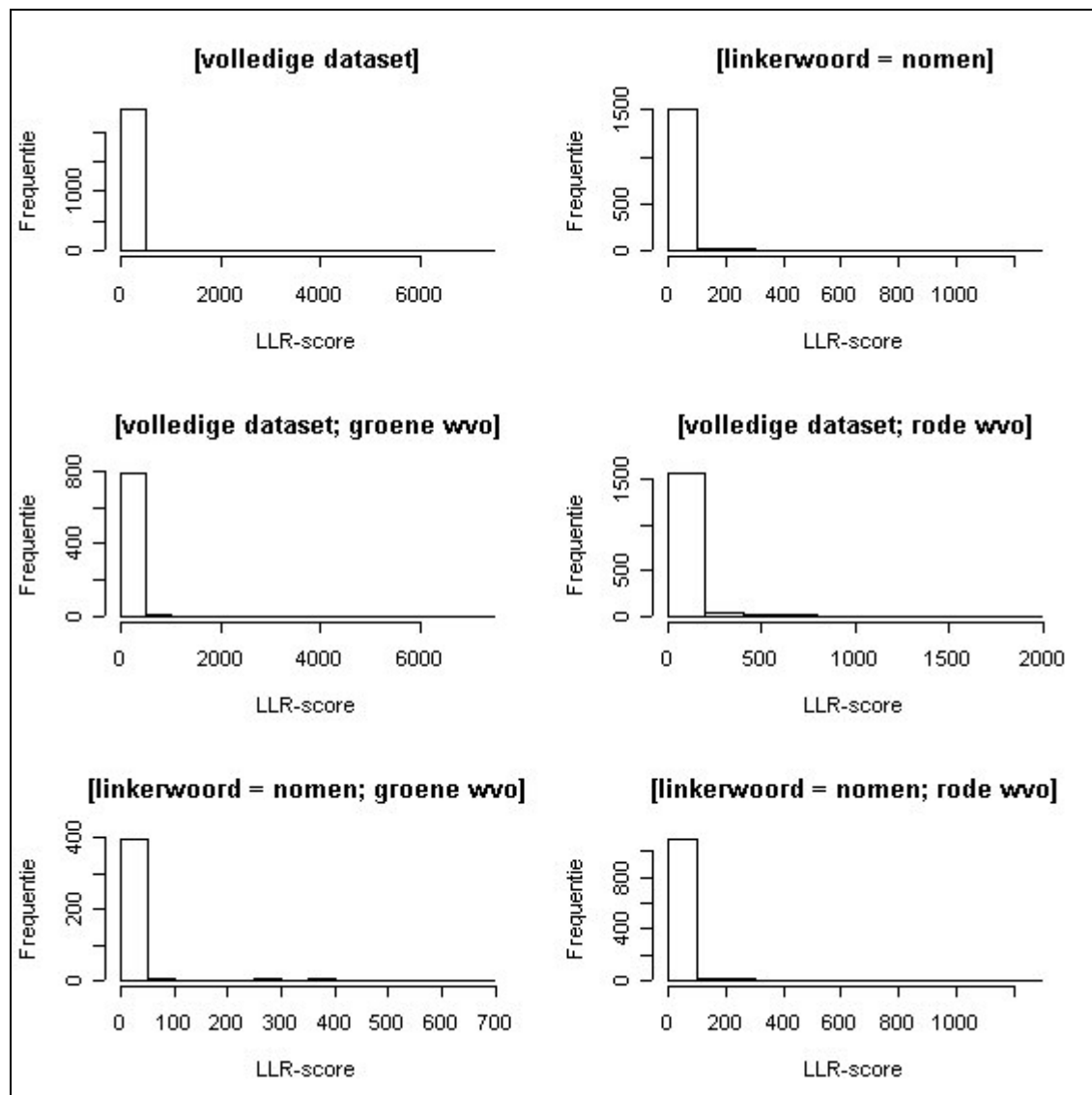
gestart met het oog op de verificatie van het accentueringsprincipe. Meer bepaald wilden we de volgende hypothese testen:

[HYP 4.6] Een direct (inherente) of primaire bepaling in de onmiddellijke linkercontext doet de voorkeur voor de rode volgorde [AUX+PART] stijgen.

De redenering achter [HYP 4.6] was dat de rode volgorde [AUX+PART], met het onbeklemtoonde hulpwerkwoord tussen het beklemtoonde deelwoord en de beklemtoonde primaire bepaling (i.c. vaste verbinding), een accentenbotsing kan vermijden.

Concreet betekent dat dat we een positieve associatie verwachten tussen de LLR-waarde—hoe hoger de LLR-waarde, hoe gefixeerder—en de rode volgorde [AUX+PART]. Omdat we hier te maken hebben met een continue variabele—bijna elke observatie heeft een unieke score—is het duidelijk dat we de gegevens niet zo makkelijk meer kunnen samenvatten in een eenvoudige tabel (zoals hierboven telkens gebeurd is).²² Om de invloed van de vaste verbinding in termen van LLR te testen, moeten we om die reden ook gebruik maken van andere types statistische technieken (met andere, meer aan een continue verdeling aangepaste, assumpties). De meest gebruikelijke verdeling voor een continue variabele is de welbekende klokvormige normale verdeling en de meest gebruikelijke statistische test die gebruikt wordt om de invloed van normaal verdeelde gegevens na te gaan, is de t-test. Omdat de betrouwbaarheid van de t-test echter staat of valt bij de vereiste dat de gegevens een normale verdeling hebben, zullen we eerst voor onze data nagaan of aan die assumptie voldaan is. In figuur 4.1 worden daarom de LLR-scores ten opzichte van de frequentie uitgezet. Op de eerste rij vinden we links de verdeling van de LLR-scores voor alle bigrammen en rechts de LLR-scores voor de bigrammen met een nominaal niet-werkwoordelijk deel ($n = 1559$). Op de tweede rij vinden we de verdeling van de LLR-scores van alle bigrammen uitgesplitst voor de groene [PART+AUX] en de rode volgorde [AUX+PART] en op de derde rij staan de verdelingen van de LLR-scores voor de bigrammen met enkel een nominaal niet-werkwoordelijk deel (verder: n-gecontroleerd), eveneens uitgesplitst naar het type woordvolgorde:

Figuur 4.1: Verdeling van de LLR-scores



Uit figuur 4.1 komt duidelijk naar voren dat de LLR-scores zeer ongelijkmatig verdeeld zijn en zich voornamelijk tussen LLR = 0 en LLR = 10 bevinden (82% van de bigrammen in de volledige dataset bevinden zich in dat interval; $n = 1960$). Dat de verdeling in de figuur bijzonder scheef is, kan ook mathematisch vastgesteld worden: voor de LLR-waarden in de volledige dataset (linksboven in figuur 4.1), bijvoorbeeld, levert die berekening een extreem hoge scheefheidswaarde op:²³ 72231.03 (ter vergelijking: een perfect symmetrische verdeling heeft een scheefheidswaarde van 0). De berekening van de scheefheid is echter slechts een descriptieve maat en zegt niets over de theoretische kans dat een dergelijke verdeling op basis van steekproefgegevens aan het toeval te wijten is en al dan niet overeenkomt met de verdeling in de volledige populatie. Met behulp van een Shapiro-Wilktest kan men dat

wel nagaan. Daaruit blijkt inderdaad dat voor geen van de 6 verdelingen in figuur 4.1 de assumptie van een normale verdeling staande gehouden kan worden (volledige dataset: $W = 0.09$; volledige dataset, groene wvo: $W = 0.06$; volledige dataset, rode wvo: $W = 0.20$; linkerwoord = nomen: $W = 0.21$; linkerwoord = nomen, groene wvo: $W = 0.17$; linkerwoord = nomen, rode wvo: $W = 0.22$; de p-waarde voor alle verdelingen is kleiner dan .0001, d.w.z. dat de nulhypothese dat de gegevens normaal verdeeld zijn verworpen moet worden).

Omdat de t-test niet gebruikt mag worden om het effect te testen van de graad van gefixeerdheid op de woordvolgorde, moeten we onze toevlucht zoeken in een statistische test voor continue gegevens die geen normaliteit veronderstelt, zoals de Wilcoxon rank sum test. De Wilcoxontest sorteert de bigrammen op basis van de LLR-waarde van groot naar klein en kent vervolgens aan elk bigram een rangnummer toe dat de positie van het bigram in het continuüm weergeeft (de grootste LLR-waarde krijgt het rangnummer 1 mee, het tweede grootste 2, enzovoort). De nulhypothese van de Wilcoxontest zegt dat het gemiddelde van alle rangnummers voor de groene volgorde [PART+AUX] niet verschilt van het gemiddelde van de bigramrangnummers voor de rode volgorde [AUX+PART]. Indien zou blijken dat het gemiddelde van de rangnummers voor een van de volgordevarianten toch wezenlijk groter of kleiner is dan de andere variant, dan kan aan de hand daarvan besloten worden dat er een verband is tussen de positie van de bigrammen in het continuüm en de keuze voor een van de volgordes. In ons geval luidt de alternatieve hypothese dat voor de groep met de rode volgorde [AUX+PART] het gemiddelde rangnummer kleiner is dan voor de groep met de groene volgorde [PART+AUX] (een kleiner gemiddelde wijst erop dat de rode bigrammen [AUX+PART] zich hogerop in de rangschikking bevinden, dus een hogere LLR-waarde hebben, en daarom als gefixeerder beschouwd kunnen worden).

Voor de volledige dataset geeft de Wilcoxontest geen significant effect aan van de mate van gefixeerdheid op woordvolgorde ($W = 649659.5$, $p > 0.05$), maar indien we enkel de n-gecontroleerde gegevens in de analyse betrekken, dan vertoont de Wilcoxontest wel een significant effect ($W = 212821$, $p = 0.002$). Hieruit blijkt dat de linguïstisch gecontroleerde statistische aanpak voor het huidige onderzoek de beste resultaten vertoont. Daarmee is niet gezegd dat die benadering de beste

operationalisering is van de factor *vaste verbinding*, de resultaten van de analyses wijzen er enkel op dat de keuze voor de woordvolgorde niet zozeer bepaald lijkt te worden door de mate waarin een werkwoord en een niet-verbaal woord dat zich net voor de eindgroep bevindt, gefixeerd zijn, maar wel door de mate waarin de combinatie van een nomen en een werkwoord gefixeerd is.

Nu is er weliswaar voor de n-gecontroleerde gegevens een significant effect vastgesteld, maar daarmee weten we nog niet in welke richting dat effect gaat (het zou kunnen dat net de groene [PART+AUX] bigrammen significant hoger gerangschikt staan). Aan de hand van een aantal centrummaten die in tabel 4.12 gepresenteerd worden, zullen we de resultaten van de beide Wilcoxontests beter trachten te begrijpen.

Tabel 4.12: Centrumgegevens van de groene [PART+AUX] en de rode [AUX+PART] LLR-scores voor de volledige en de n-gecontroleerde dataset.

	Volledige dataset		Linkerwoord = nomen	
	LLR-score	LLR-score	LLR-score	LLR-score
	Groene wvo	Rode wvo	Groene wvo	Rode wvo
Gemiddelde rangnummers	1015.01	1046.984	1259.056	1123.576
Mediaan	0.029	0.016	0.0050	0.0060
Gemiddelde	26.83	21.86	10.95	18.53

In de eerste rij van tabel 4.12 worden de gemiddelde LLR-rangnummers voor de groene [PART+AUX] en rode [AUX+PART] woordvolgorde in de volledige en in de n-gecontroleerde dataset gegeven. Daar is te zien dat het verschil tussen de beide woordvolgordes in de volledige dataset veel kleiner is dan in de n-gecontroleerde dataset ($|1015.01-1046.984| < |1259.056-1123.576|$), hetgeen verklaart waarom de analyse enkel op die laatste dataset een significant effect vertoonde. Opvallend daarbij is dat de gemiddelde rangorde voor de rode bigrammen [AUX+PART] in de volledige dataset groter is dan de groene volgorde [PART+AUX], wat erop wijst dat de

bigrammen in de rode volgorde [AUX+PART], niet gecontroleerd voor woordsoort, gemiddeld genomen een lagere LLR-waarde hebben, en dus minder gefixeerd zijn.

In de gecontroleerde dataset ligt de verhouding net omgekeerd. Daaruit kunnen we afleiden dat de bigrammen met een nominaal niet-werkwoordelijk deel die een lagere LLR-waarde hebben (en dus meer gefixeerd zijn) vooral in de rode volgorde [AUX+PART] verschijnen (bevestiging van [HYP 4.6]). Eenvoudiger geformuleerd: een vaste(re) verbinding met een nominaal preverbaal deel doet de voorkeur voor de rode volgorde [AUX+PART] stijgen.

Als de gemiddelde rangnummers voor de verschillende niet-nominale woordsoorten uitgesplitst zouden worden, dan zou men inderdaad kunnen zien dat de groene [PART+AUX] bigrammen gemiddeld genomen een hogere rang bekleden, met uitzondering van de bigrammen met een prepositie (groen: rang 197, rood: rang 162): voor de bigrammen met een pronomen in de groene volgorde [PART+AUX] is dat rang 604 tegenover rang 703 voor die in de rode volgorde [AUX+PART], voor de bigrammen met een adjectief of adverbium is dat rang 831 als ze in de groene volgorde [PART+AUX] staan en rang 897 als ze in de rode volgorde [AUX+PART] staan (de bigrammen met een conjunctie staan gemiddeld op rang 785 in de groene volgorde [PART+AUX]; er zijn geen bigrammen met een conjunctie in de rode volgorde [AUX+PART], zodat ook geen rangnummer berekend kon worden). De andere centrumgegevens, de gemiddelde LLR-waarde en de mediaan van de LLR-waarden, geven dezelfde tendensen weer (merk op dat een hogere LLR-score correleert met een grotere mate van gefixeerdheid).

We sluiten de discussie naar het effect van de vaste verbinding af met een vertrouwde chi-kwadraatanalyse door voor de volledige dataset de bigrammen onder te verdelen in een groep met de significante bigrammen en groep met de niet-significante bigrammen.

Tabel 4.13: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type bigram.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Bigram = niet significant	34.46% (622/1805)	65.54% (1183/1805)
Bigram = significant	28.55% (167/585)	71.45% (418/585)

Anders dan hierboven het geval was, komt uit tabel 4.13 naar voren dat voor de volledige dataset, conform de hypothese, de keuze voor de groene volgorde [PART+AUX] iets groter is als het bigram niet-significant (niet gefixeerd) is dan als het dat wel is, maar een frappant verschil is er zeker niet. Desondanks geeft de statistische analyse aan dat het verschil tussen de beide categorieën significant is ($\chi^2 = 6.98$, $df = 1$, $p < .01$); de o.r. van 1.31 (B.I. = 1.07-1.61) wijst erop dat de kans op een groene volgorde [PART+AUX] 31% groter is als het bigram niet significant (niet gefixeerd) is dan als het wel significant is. Daaruit kunnen we dan net als hierboven concluderen dat de aanwezigheid van een niet-werkwoordelijk deel dat een eenheid vormt met het werkwoord de kans op een rode volgorde [AUX+PART] doet stijgen (bevestiging van [HYP 4.6]).

Indien we enkel zouden kijken naar de n-gecontroleerde gegevens, dan is het onderscheid tussen significante en niet-significante bigrammen nog significanter ($\chi^2 = 25.47$, $df = 1$, $p < .0001$) en is de o.r.-waarde nog extremer: de kans op een groene volgorde [PART+AUX], in vergelijking met de kans op een rode volgorde [AUX+PART], is 2.25 keer groter indien het bigram niet significant is (o.r. = 2.25, B.I. = 1.62-3.16).

Tabel 4.14: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type bigram voor de n-gecontroleerde dataset.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Bigram = niet significant	29.45% (361/1226)	70.55% (865/1226)
Bigram = significant	15.66% (52/332)	84.34% (280/332)

In de voorgaande analyses hebben we op overtuigende wijze evidentie gevonden voor het effect van de vaste verbinding op de keuze voor de groene [PART+AUX] of rode [AUX+PART] volgorde. Uit drie van de vier analyses is gebleken dat de aanwezigheid van een vaste verbinding, geoperationaliseerd in termen van LLR-waarden, de voorkeur voor de rode volgorde [AUX+PART] doet stijgen. Dat resultaat strookt perfect met het accentueringsprincipe dat in paragraaf 4.1 voorgesteld werd, aangezien de rode volgorde [AUX+PART], met het onbeklemtoonde hulpwerkwoord tussen het preverbale accent en het deelwoordaccent, een accentenbotsing kan vermijden. Daarbij moet wel aangenomen worden dat (i) het preverbale deel van een vaste verbinding een zwaar accent draagt en (ii) dat de afstand tussen de beklemtoonde syllabes van het deelwoord en dat preverbale deel minder dan 2-3 syllabes bedraagt. In paragraaf 4.3.3.6 zullen we aan de hand van de gerichtere prosodische analyse uit paragraaf 4.3.3.5 nagaan in hoeverre die veronderstellingen correct zijn.

Op methodologisch en analytisch vlak hebben we kunnen vaststellen dat de operationalisering van de factor *vaste verbinding* in termen van LLR-waarden een aantal onmiskenbare voordelen heeft: niet alleen is de operationalisering een objectieve manier om voor reële taalgegevens te bepalen of ze al dan niet een vaste verbinding zijn (wat met behulp van traditionele linguïstische criteria alleen schier onmogelijk is), ze heeft ook het voordeel dat de gegevens op een schaal van minder naar meer gefixeerd geplaatst kunnen worden en dat daarom krachtigere statistische tests gebruikt kunnen worden om naar associaties te zoeken en, niet onbelangrijk, ze levert plausibele resultaten op.

Toch moet dat ons er niet van weerhouden om een aantal kritische kanttekeningen te plaatsen bij het gebruik van de LLR-test. Zo is er hierboven al op gewezen dat de kans reëel is dat een aantal vaste verbindingen door de mazen van het net geglipt zijn doordat enkel rekening gehouden is met het onmiddellijke linkerwoord van de werkwoordelijke eindgroep. In toekomstig onderzoek zal daarom zeker nagegaan moeten worden wat het effect van die restrictieve procedure geweest is. Toekomstig onderzoek zal zich verder ook moeten buigen over de vraag wat een vaste verbinding tot een vaste verbinding maakt en of de beslissing in sommige van de analyses om enkel de combinaties met een nomen als niet-werkwoordelijk deel te bestuderen, wel gerechtvaardigd was.

4.3.3.4 *Lengte van het middenstuk*

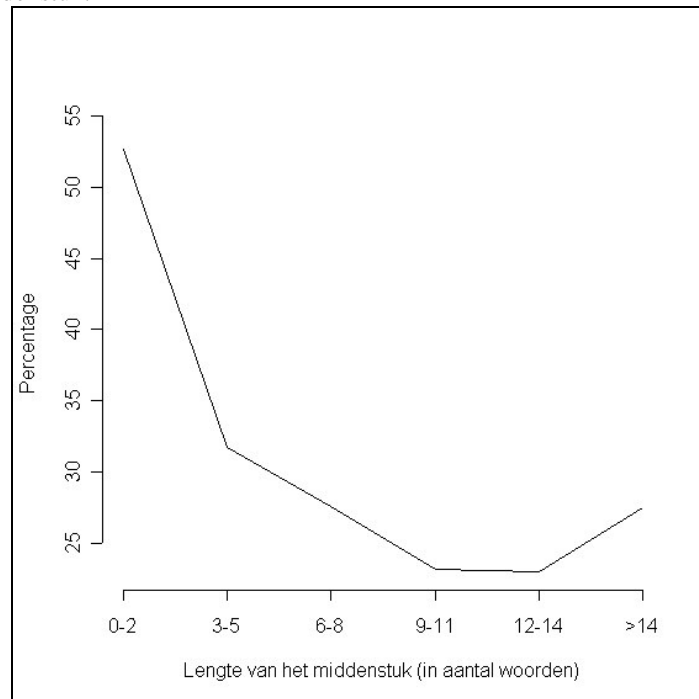
Een laatste indirecte operationalisering van de invloed van de accentverdeling in de linkercontext op de woordvolgorde in de tweeledige eindgroep is de lengte van het middenstuk, opgevat als het aantal woorden tussen het grammaticaal verbindend voegwoord *dat* en de werkwoordelijke eindgroep. We herhalen hier de hypothese, gebaseerd op de bevindingen van eerder wetenschappelijk onderzoek (zie paragraaf 4.3.1):

[HYP 4.7] Naarmate de lengte van het middenstuk toeneemt, des te meer rode volgordes [AUX+PART] gekozen worden.

De redenering achter [HYP 4.7] is dat de keuze voor de rode volgorde [AUX+PART], met het hulpwerkwoord tussen het preverbale accent en het deelwoordaccent, een accentenbotsing kan vermijden (daarbij wordt aangenomen dat de lengte van het middenstuk positief correleert met de kans dat een sterk beklemtoond element voor de werkwoordelijke eindgroep komt te staan).

Om na te gaan of de resultaten uit eerder onderzoek bevestigd kunnen worden, hebben we in figuur 4.2 de voorkeur voor de groene volgorde [PART+AUX] uitgezet tegenover de lengte van het middenstuk (in termen van woorden).

Figuur 4.2: Relatieve proporties van de groene [PART+AUX] woordvolgorde als een functie van de lengte van het middenstuk.



De resultaten in figuur 4.2 laten, conform de resultaten van eerder onderzoek, een graduele afname zien van de groene volgorde [PART+AUX] naarmate de lengte van het middenstuk groter wordt. Die afname is bijzonder opvallend als we de overgang van de zinnen met een middenstuk van 0 tot 2 woorden naar de zinnen met een middenstuk van 3 tot 5 woorden bekijken. Wat achter die spectaculaire daling zou kunnen zitten, is niet zo gauw uit te maken.

Verder kan nog vastgesteld worden dat de graduele afname geen perfect lineaire trend is, aangezien de zinnen met een erg lang middenstuk (meer dan 14 woorden) opnieuw een wat grotere voorkeur voor de groene volgorde [PART+AUX] hebben dan zinnen met een middenstuk van 9 tot 14 woorden. Ook hier is niet onmiddellijk duidelijk wat die plotselinge stijging veroorzaakt heeft. In ieder geval is het zo dat, op voorwaarde dat we mogen aannemen dat er inderdaad een positieve associatie is tussen de lengte van het middenstuk en de kans dat een beklemtoond element net voor de werkwoordelijke eindgroep verschijnt, de gegevens in figuur 4.2 de accentueringshypothese lijken te ondersteunen: hoe langer het middenstuk en dus hoe groter de kans op een accent net voor de werkwoordelijke eindgroep, hoe groter de voorkeur voor de rode volgorde [AUX+PART] die in een dergelijke context een

accentenbotsing kan vermijden ($MH-\chi^2 = 64.38$, $df = 1$, $p < .0001$; $\gamma = 0.28$, $ASE = 0.03$; bevestiging van [HYP 4.7]).

Hoe plausibel de verklaring van het vastgestelde effect ook moge lijken, het is duidelijk dat de assumptie dat een langer middenstuk een grotere kans heeft op een zwaar beklemtoond preverbaal element in verder onderzoek geverifieerd moet worden. Daarenboven kan ook hier tegen de gebruikte operationalisering ingebracht worden dat ze de accentueringshypothese op een syntactische manier test en niet op een prosodische manier: ook al bevestigt het resultaat in de bovenstaande grafiek de resultaten van de vorige onderzoeken en lijkt het er met andere woorden op dat een accentueringsprincipe werkzaam is bij de keuze van woordvolgorde, toch moet een meer gerichte operationalisering uitmaken of er inderdaad een invloed is van de accenten binnen de linkerconstituent. De volgende paragraaf komt tegemoet aan die eis.

4.3.3.5 *Afstand van het deelwoordaccent tot het laatste preverbale woordaccent*

Net als bij de verificatie van het accentueringsprincipe binnen de werkwoordelijke eindgroep (4.3.1) moeten we ook hier de vraag stellen in hoeverre de operationalisering in termen van definietheid en inherentie van het preverbale taal materiaal en lengte van het middenstuk accuraat is. We brengen daarbij in herinnering dat het accentueringsprincipe geconcipieerd is als een relatief correctiemechanisme dat enerzijds het aantal zwaar beklemtoonde lettergrepen uit elkaar tracht te houden en anderzijds ervoor tracht te zorgen dat het aantal onbeklemtoonde lettergrepen tussen twee beklemtoonde lettergrepen niet al te groot wordt (2-3 lettergrepen). Vanuit die optiek zou men de syntactico-pragmatische operationalisering uit de vorige paragrafen veeleer grofkorrelig kunnen noemen, omdat ze er enkel rekening mee houden of een woord, woordgroep, constituent of reeks constituenten een accent draagt en niet welke syllabe van dat woord, die woordgroep, constituent of reeks constituenten het accent draagt.

Om tegemoet te komen aan die bedenking en het accentueringsprincipe op een meer fijnkorrelige manier te verifiëren, werd het aantal syllabes tussen het woordaccent van het deelwoord en het woordaccent van het laatste preverbale woord geteld. Indien de enige klinker (nucleus) van het laatste preverbale woord een sjwa is,

dan werd daarvan aangenomen dat het geen accent kan dragen (en daarom ook geen prosodisch woord op zichzelf kan vormen; Booij 1995) en werd de afstand tussen het woordaccent van het deelwoord en het woordaccent van het voorlaatste preverbale woord geteld. De identificatie van de woordaccenten gebeurde in eerste instantie aan de hand van de prosodische informatie in EGVD 1.4 (2004) en, indien het een eigennaam betrof, de taaldatabanken van de VRT (<http://www.taaldatabanken.com/>). Als het laatste preverbale woord een samenstelling is die niet in EGVD 1.4 (2004) opgenomen is, werd ervan uitgegaan dat het primaire accent op het eerste lid van de samenstelling komt te liggen (*AFluisteràpparaat*).²⁴ Van woorden die een wisselend accent hebben, werd in EGVD 1.4 (2004) opgezocht op basis van welke parameters de accentlocatie bepaald kon worden. Bij een registersverschil werd gekozen voor de standaardtalige, niet-Belgische variant. Afkortingen werden tot slot als monosyllabisch opgevat (met dus één woordaccent en geen daaropvolgende onbeklemtoonde lettergrepen).

Bij de berekening van de afstand tussen de twee woordaccenten werd opnieuw, zoals in paragraaf 4.2, het aantal syllabes van het hulpwerkwoord niet meegeteld, om een incorporatie van de responsvariabele in de verklarende variabele te vermijden.

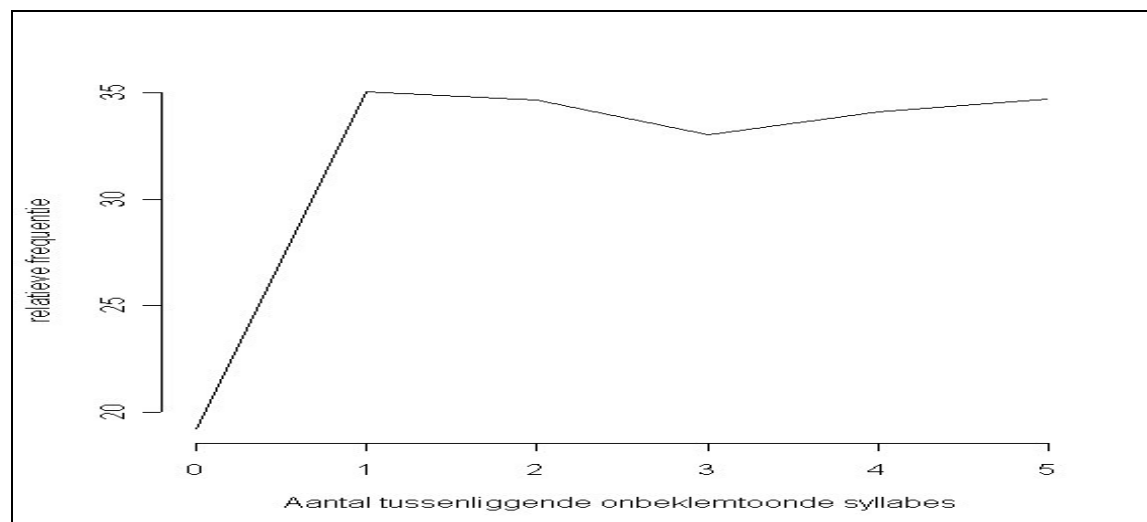
In figuur 4.3 worden de gegevens die in de groene [PART+AUX] en de rode volgorde [AUX+PART] voorkomen onderverdeeld in verschillende categorieën van een gradueel toenemende afstand tussen de woordaccenten van het laatste preverbale woord en het deelwoord. De hypothese daarbij is de volgende:

[HYP 4.8] Naarmate de afstand tussen het laatste accent in de linkercontext en het begin van de werkwoordelijke eindgroep kleiner wordt, worden meer rode woordvolgordes [AUX+PART] geprefereerd.

Indien nl. het woordaccent van het laatste preverbale woord zich op minder dan 2-3 syllabes van het deelwoordaccent bevindt (het hulpwerkwoord buiten beschouwing gelaten), dan wordt de rode volgorde [AUX+PART] (met het hulpwerkwoord tussen het laatste preverbale woord en het deelwoord) geprefereerd om zo een accentenbotsing te vermijden. Omgekeerd zal de groene volgorde [PART+AUX] de voorkeur wegdragen indien de afstand tussen de twee accenten te groot dreigt te worden. In figuur 4.3 zetten we het percentage groene woordvolgordes [PART+AUX]

uit ten opzichte van het aantal tussenliggende onbeklemtoonde syllabes tussen het woordaccent van het deelwoord en het woordaccent van het laatste beklemtoonbare preverbale woord:

Figuur 4.3: Relatieve proporties van de groene [PART+AUX] woordvolgorde per afstand van het deelwoordaccent tot het laatste preverbale woordaccent (in termen van tussenliggende onbeklemtoonde syllabes).



Zoals uit de figuur duidelijk wordt, is vooral het verschil tussen 0 en meer dan 0 tussenliggende onbeklemtoonde syllabes opmerkelijk: terwijl er bij 0 tussenliggende syllabes zeer gering gebruik gemaakt wordt van de groene woordvolgorde [PART+AUX] (19.17%, $n = 46/240$), stijgt de voorkeur voor de groene volgorde [PART+AUX] naarmate de afstand tussen de twee woordaccenten groter wordt (gemiddeld: 34.56%; $n = 743/2150$). De lineaire trend is statistisch significant (MH- $\chi^2 = 4.09$, $df = 1$, $p < .05$; $\gamma = -0.07$, $ASE = 0.03$) en wijst erop dat met een grote mate van zekerheid geponeerd kan worden dat de keuze voor de rode volgorde [AUX+PART] groter is wanneer de twee woordaccenten onmiddellijk met elkaar in botsing komen. Conform de accentueringshypothese is de voorkeur voor de rode volgorde [AUX+PART] in die omstandigheden groter, omdat het tussenliggende onbeklemtoonde hulpwerkwoord tussen de beide accenten geplaatst wordt en zo de accentenbotsing vermeden kan worden²⁵ (bevestiging van [HYP 4.8]). Tegen de achtergrond van het principe blijft het echter wel vreemd dat er blijkbaar geen enkel probleem is in het reële taalgebruik om 3 of meer onbeklemtoonde syllabes te hebben zonder dat de keuze voor de groene volgorde [PART+AUX] verder stijgt (zoals in paragraaf 4.2 al vastgesteld werd).

4.3.3.6 *Vergelijking van de verschillende operationaliseringen*

In de vorige paragrafen hebben we middels diverse operationaliseringen getracht om de impact van de accentverdeling in de linkercontext op de woordvolgorde te analyseren. Daarbij zijn de definietheid van de laatste substantiefgroep, de informationaliteit van het laatste linkerwoord, de inherentie van de laatste preverbale constituent, de lengte van het middenstuk en de afstand tussen het laatste preverbale woordaccent en het deelwoordaccent aan bod gekomen. De resultaten van de analyses waren op zijn zachtst gezegd bemoedigend, in die zin dat over het algemeen de vooropgestelde hypothesen bevestigd werden. De resultaten gaven ook aan dat er een accentueringsprincipe werkzaam lijkt te zijn dat ervoor zorgt dat beklemtoonde syllabes niet in elkaars onmiddellijke buurt voorkomen en dat de zin afgesloten wordt met een lage grenstoon.

Ondanks dat succes moeten we toch behoedzaam omspringen met de interpretatie van de resultaten in termen van accentuering. We hebben er in dit hoofdstuk nl. al meermaals op gewezen dat de indirecte (syntactico-pragmatische) operationaliseringen uit de paragrafen 4.3.3.1 t.e.m. 4.3.3.4 veeleer grofkorrelig zijn, omdat ze er enkel rekening mee houden of een woord, woordgroep, constituent of reeks constituenten een accent draagt en niet welke syllabe van dat woord, die woordgroep, constituent of reeks constituenten het accent draagt. In het kader van een principe dat expliciet uitspraken doet over ideale en minder ideale accentpatronen is dat geen onbelangrijk detail: een extra onbeklemtoonde syllabe tussen het deelwoordaccent en het laatste preverbale accent kan de voorkeur voor de rode volgorde [AUX+PART] nl. drastisch doen dalen (omdat de kans op een accentenbotsing daalt).

In deze paragraaf willen we nagaan in hoeverre die kritische houding terecht is. Daartoe zullen we de effecten van de verschillende indirecte (syntactico-pragmatische) operationaliseringen vergelijken met de meest directe operationalisering, i.e. de operationalisering in termen van beklemtoonde en onbeklemtoonde syllabes (paragraaf 4.3.3.5). We gebruiken daarvoor de in hoofdstuk 3 voorgestelde CMH-statistiek, waarbij het effect van een verklarende variabele op de responsvariabele getest wordt onder controle van een tweede verklarende variabele. Toegepast op de analyse die we hier willen uitvoeren, betekent dat dat we zullen nagaan wat het effect is van de

verschillende indirecte operationaliseringen op de woordvolgorde onder controle van de directe operationalisering. De verwachting is de volgende: indien het inderdaad zo is dat de verschillende indirecte operationaliseringen geïnterpreteerd mogen worden in prosodische termen, dan verwachten we dat het effect van die operationaliseringen gelijkloopt met dat van de directe operationalisering, waardoor het effect van de indirecte operationaliseringen in de CMH-analyse opgeslokt zal worden (i.e. zal verdwijnen) door de directe operationalisering.

Uit de CMH-analyse waar het effect van informationaliteit onder controle van de accentuering onderzocht is²⁶, blijkt alvast dat er geen sprake kan zijn van een dergelijk opslokkend effect: het effect van informationaliteit blijft nl. onverminderd bestaan, niet alleen in de marginale tabel (CMH- $\chi^2 = 118.85$, $df = 2$, $p < .0001$), maar ook in de verschillende partiële tabellen (interval = 0-1 syll.: $\chi^2 = 80.92$, $df = 2$, $p < .0001$; interval = 2-3 syll.: $\chi^2 = 33.76$, $df = 2$, $p < .0001$; interval > 3 syll.: $\chi^2 = 7.85$, $df = 2$, $p < .02$).

De CMH-analyse van het effect van gefixeerdheid (in termen van niet-gecontroleerde bigrammen) onder controle van accentuering levert een diffuser beeld op: alhoewel de globale CMH-statistieken op een significant effect van gefixeerdheid wijzen (CMH- $\chi^2 = 6.25$, $df = 1$, $p < .02$), geeft de Breslow-Day-statistiek dat er geen homogeniteit bestaat onder de o.r.'s in de partiële tabellen (Br-D- $\chi^2 = 7.65$, $df = 2$, $p < .03$). Bij nadere bestudering van de partiële tabellen blijkt er een interactie-effect te zijn: terwijl het verschil tussen de significante en niet-significante bigrammen niet significant blijkt te zijn voor de gevallen met een minimaal onbeklemtoond interval ($\chi^2 = 0.39$, $df = 1$, $p > .05$), is er wel een significant verschil tussen de beide types voor de gevallen waarbij het interval groter is (interval = 2-3 syll.: $\chi^2 = 5.56$, $df = 1$, $p < .02$; interval > 3 syll.: $\chi^2 = 7.39$, $df = 1$, $p < .007$). Daaruit blijkt in ieder geval dat het effect van gefixeerdheid niet helemaal gelijkloopt met het effect van accentuering.

De CMH-analyse waarin het effect van de lengte van het middenstuk onder controle van de accentdistributie bestudeerd wordt, levert eveneens een globale significante CMH-statistiek op (CMH- $\chi^2 = 101.54$, $df = 5$, $p < .0001$), waardoor ook hier geen sprake kan zijn van een volledige identiteit tussen de variabelen *lengte van het middenstuk* en *accentdistributie in de linkercontext*. De bestudering van de partiële

tabellen brengt eveneens een interactie-effect aan het licht: de lengte van het middenstuk blijkt nl. enkel significant te zijn wanneer de afstand tussen het vorige accent en het deelwoord accent minimaal of intermediair is (interval = 0-1 syll.: $MH-\chi^2 = 39.54$, $df = 1$, $p < .0001$; interval = 2-3 syll.: $MH-\chi^2 = 24.97$, $df = 1$, $p < .0001$; interval > 3 syll.: $MH-\chi^2 = 1.76$, $df = 1$, $p > .05$).

Ondanks het feit dat de resultaten van de CMH-analyses erop wijzen dat de verschillende indirecte operationaliseringen niet geheel onafhankelijk zijn van de accentdistributie in hun effect op de woordvolgorde, is het toch duidelijk geworden dat geen van de operationaliseringen er volledig mee samenvalt. Dat lijkt er met andere woorden op te wijzen dat we die effecten (gedeeltelijk) op een andere manier zullen moeten interpreteren dan in termen van accentuering. In hoofdstuk 7 zullen we daartoe een poging ondernemen.

4.4 De impact van de accentverdeling in de rechtercontext

Nu de invloed van de accentverdeling in de werkwoordelijke eindgroep en links van de werkwoordelijke eindgroep duidelijk geworden is, zal in een laatste stap de invloed van de accentuering in de rechtercontext van de eindgroep op de woordvolgorde bestudeerd worden. Na een overzicht van de belangrijkste literatuur ter zake (paragraaf 4.4.1) en de hypothesen die aan de grondslag zullen liggen van ons onderzoek (paragraaf 4.4.2), bespreken en evalueren we de resultaten (paragraaf 4.4.3). Conform het aantal operationaliseringen zullen de resultaten in drie luiken gepresenteerd worden: paragraaf 4.4.3.1 is gewijd aan het effect van de aan- vs. afwezigheid van een laatste zinsplaats, paragraaf 4.4.3.2 zoomt in op de impact van de grammaticale relatie tussen de laatste zinsplaats en het hoofd en in paragraaf 4.4.3.3 wordt de afstand van het deelwoordaccent tot het eerste postverbale woordaccent besproken. In paragraaf 4.4.3.4 bieden we tot slot een korte, tussentijdse evaluatie van de resultaten in deze paragraaf, als opstapje naar de globale vergelijking van de accentdistributie in, voor en na de werkwoordelijke eindgroep in paragraaf 4.5.

4.4.1 *Status quaestionis*

Net als de studie naar de invloed van de linkercontext totnogtoe nooit aan de hand van prosodische factoren onderzocht werd (met uitzondering van Swerts 1998), maar veeleer aan de hand van syntactico-pragmatische factoren, is ook het effect van de rechtercontext nooit prosodisch onderzocht. In het geval van de rechtercontext werd enkel gekeken of de aan- of afwezigheid van een laatste zinsplaats (ANS 1997: 1364-1386) een invloed had op de keuze voor de woordvolgorde.²⁷ De studie van het effect van de laatste zinsplaats op de woordvolgorde in de werkwoordelijke eindgroep is echter nooit echt van centraal belang geweest in het onderzoeksdomein. De belangrijkste reden daarvoor is ongetwijfeld dat er nooit eensgezindheid geweest is over de vraag of de laatste zinsplaats al dan niet een invloed uitoefent op de woordvolgorde. Zo is De Schutter (1976, 1996) de enige onderzoeker die een statistisch significant effect heeft kunnen vaststellen tussen de aanwezigheid vs. afwezigheid van een laatste zinsplaats en de keuze voor een van de woordvolgordes (De Cubber 1986, Haeseryn 1990, Jansen 1990, Swerts 1998 vonden geen effect van de laatste zinsplaats). De resultaten van dat onderzoek, die zowel betrekking hebben op gesproken als geschreven taal (althans voor het Nederlands-Nederlandse materiaal dat De Schutter bestudeert), maken duidelijk dat de aanwezigheid van een laatste zinsplaats de kans op een groene volgorde [PART+AUX] doet stijgen. De Schutter (1996: 210) merkt daarbij op dat voor het effect van de laatste zinsplaats in de woordvolgordebepaling zeker geen hoofdrol weggelegd is, maar hij meent toch dat het effect geïnterpreteerd kan worden in de lijn van het accentueringsprincipe. Zoals gezegd bestaat dat principe uit twee componenten, m.n. de evenwichtige verdeling van zinsaccenten en het streven naar het behoud van het platte-hoedpatroon. De grotere voorkeur voor de groene volgorde [PART+AUX] in zinnen met een laatste zinsplaats kan men nu, zo meent De Schutter, begrijpen als de neiging om een accentenbotsing te vermijden: doordat het ongeaccentueerde hulpwerkwoord het zinsaccent op de laatste zinsplaats van het secundaire accent op het verleden deelwoord voldoende (twee tot drie lettergrepen) gescheiden houdt, kan een accentenbotsing vermeden worden. Daarbij moet wel aangenomen worden dat postverbale, ‘geëxtraponeerde’ constituenten in focus staan en daardoor een hoofdaccent dragen (zie ook Van Donselaar 1994, Selkirk 1995, Nooteboom 1997).

- (16a) [...] dat het gepraat had met bronnen dichtbij Dole en Kemp en daar het nieuws vernomen had. (DS)
- (16b) [...] dat het had gepraat met bronnen dichtbij Dole en Kemp en daar het nieuws vernomen had.

In zin (16a) draagt de groene volgorde [PART+AUX] de voorkeur, omdat het ervoor zorgt dat de hoofdaccenten op *gePRAAT* en *BRonnen* ver genoeg van elkaar verwijderd zijn: de twee ongeaccentueerde lettergrepen op *had* en *met* komen tussen de beide geaccentueerde lettergrepen in te staan, waardoor er een interval van twee ongeaccentueerde lettergrepen ontstaat. Indien de rode volgorde [AUX+PART] gekozen was (voorbeeldzin (16b)), dan zouden de geaccentueerde lettergrepen slechts door één enkele ongeaccentueerde lettergreep gescheiden worden, m.n. door *met*, waardoor de twee sterk geaccentueerde lettergrepen te snel op elkaar zouden volgen.

4.4.2 Hypotheses

Op basis van de net besproken literatuur kunnen we de volgende hypothese extraheren:

- [HYP 4.9] De aanwezigheid (versus afwezigheid) van een laatste zinsplaats doet de voorkeur voor de groene volgorde [PART+AUX] stijgen (paragraaf 4.4.3.1).

Naast [HYP 4.9] zullen we ook een wat specifiekere (syntactische) hypothese testen, waarbij we de globale categorie *laatste zinsplaats* differentiëren naar grammaticale relatie (i.c. het verschil tussen adjunct en complement). Alhoewel de invloed van de grammaticale relatie op de keuze voor de woordvolgorde in de werkwoordelijke eindgroep nog nooit onderzocht is, hebben een aantal wetenschappers in ander onderzoek de hypothese geuit dat de grammaticale relatie tussen een constituent en een hoofd niet alleen iets zegt over syntactische of pragmatische verhoudingen, maar ook over de beklemtoning. Zo meent Braecke (1990: 129-130) dat complementen, in tegenstelling tot adjuncten, meer geschikt zijn om in focus te staan en daardoor vaker beklemtoond worden.²⁸ Als dat inderdaad zo zou zijn, dan kunnen we verwachten dat postverbale complementen vaker door een groene volgorde [PART+AUX]

voorafgegaan zullen worden dan postverbale adjuncten, omdat de groene volgorde [PART+AUX], met het onbeklemtoonde hulpwerkwoord tussen het deelwoordaccent en het postverbale accent, een accentenbotsing kan vermijden:

[HYP 4.10] Een postverbaal complement (versus een postverbaal adjunct) doet de voorkeur voor de groene volgorde [PART+AUX] stijgen (paragraaf 4.4.3.2).

Naast de indirecte, syntactische operationalisering van de accentverdeling in de linkercontext, zullen we in deze sectie, net zoals in de vorige secties (paragraaf 4.2 en 4.3), ook een meer fijnkorrelige, specifiek prosodische analyse uitvoeren, waarbij het aantal onbeklemtoonde syllabes tussen het deelwoordaccent en het accent van het eerste postverbale woord geteld zal worden. De hypothese is de volgende (gebaseerd op het accentueringsprincipe; voor meer uitleg: zie paragraaf 4.1):

[HYP 4.11] Naarmate de afstand tussen het deelwoordaccent en het eerste postverbale accent kleiner wordt, worden meer groene woordvolgordes [PART+AUX] geprefereerd (paragraaf 4.4.3.3).

De redenering achter [HYP 4.11] is dat de groene volgorde [AUX+PART], met het onbeklemtoonde hulpwerkwoord aan het eind van de eindgroep, een accentenbotsing tussen het zinsaccent op het deelwoord en het postverbaal zinsaccent kan vermijden (vergelijk voorbeeldzin (16a) en (16b)).

4.4.3 Resultaten en discussie

4.4.3.1 Aan- vs. afwezigheid van een laatste zinsplaats

We beginnen onze onderzoeken naar de impact van de accentverdeling in de rechtercontext van de werkwoordelijke eindgroep met [HYP 4.9]:

[HYP 4.9] De aanwezigheid (versus afwezigheid) van een laatste zinsplaats doet de voorkeur voor de groene volgorde [PART+AUX] stijgen.

Alvorens de kwantitatieve analyses te presenteren, zullen we eerst kort ingaan op de manier waarop uitgemaakt is of een postverbale constituent of postverbale bijzin

(verder: PVCB) een laatste zinsplaats is. Een PVCB is daarbij gedefinieerd als ‘alles wat zich tussen de werkwoordelijke eindgroep en een zinsafsluitend interpunctieteken (punt, uitroepeteken of vraagteken) bevindt’. Echter, niet alle PVCB's zijn laatste zinsplaatsen; om als laatste zinsplaats te kunnen gelden, moet de PVCB nl. een syntactische functie hebben binnen de zin waarvan het deelwoord het hoofdwerkwoord is. Voorbeeldzin (17) geeft een voorbeeld van een PVCB die geen laatste zinsplaats is.

- (17) [...] dat hij in Harrisburg meteen zijn vrouw *had gebeld* en haar had verteld dat zij en de kinderen het mooiste zijn wat hem in zijn leven is overkomen. (DS)

De onderstreepte PVCB in voorbeeldzin (17) is nevenschikt aan de bijzin met de targeteindgroep (gecursiveerd) en vervult derhalve geen functie in de targetbijzin, waardoor de PVCB niet als laatste zinsplaats beschouwd kan worden (in de hieronder gepresenteerde analyse zullen dergelijke zinsexterne PVCB's samengenomen worden met de groep zonder PVCB).²⁹ Een PVCB die wel als laatste zinsplaats fungeert kan de vorm aannemen van een prepositionele constituent, een bepaling van gesteldheid, een constituent ingeleid door een voegwoord van vergelijking of een postverbale ondergeschikte bijzin (er werden geen postverbale nominale constituenten of nevenschikte zinsdeelstukken gevonden in de dataset). De laatste zinsplaats kan daarbij uit een constituent of bijzin bestaan, waarvan een of meer andere bijzinnen afhangen:

- (18) [...] dat piloot Leul Abate tevergeefs had geprobeerd de kapers duidelijk te maken dat hij niet voldoende brandstof had om hun eis - doorvliegen naar Australië - in te willigen. (DS)

Indien de laatste zinsplaats uit twee nevenschikte constituenten bestaat, werd enkel de eerste PVCB in de analyse betrokken. Om uit te maken of een laatste zinsplaats uit een of meerdere constituenten bestaat, werd getracht een deel van de laatste zinsplaats voorop te plaatsen en het andere deel achterop te laten staan. Indien die test slaagde, werd de laatste zinsplaats als meerledig beschouwd, indien de test niet slaagde, werd de laatste zinsplaats als eenledig beschouwd. Zo bevat voorbeeldzin (19) meer dan een PVCB, voorbeeldzin (20) is eenledig:

- (19) [...] dat de Palestijnse politie de afgelopen dagen op grote schaal arrestaties heeft verricht in de hoogste rangen van het Palestijnse zelfbestuur na een mislukte staatsgreep tegen Arafat. (DS)
- (20) [...] dat de Vlaamse regering heeft geknoeid bij de aanstelling van Marc Clémeur tot intendant van de Vlopera. (DS)

Een meerledige PVCB, zoals in (19), hoeft uiteraard ook niet uit twee zinsinterne PVCB te bestaan, maar kan ook uit een zinsinterne en zinsexterne constituent/bijzin opgebouwd zijn. Het spreekt voor zich dat voor de verdere analyses enkel rekening gehouden werd met wat traditioneel als laatste zinsplaats begrepen wordt.³⁰

Nu duidelijk is wat begrepen wordt als een laatste zinsplaats en wat niet, kunnen we de resultaten van vorig onderzoek (m.n. dat van De Schutter) verifiëren. Daartoe worden in tabel 4.15 de gegevens in onze dataset verspreid over de twee volgordealternatieven en de categorieën van zinnen met en zinnen zonder laatste zinsplaats:

Tabel 4.15: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde met en zonder een laatste zinsplaats.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Zonder laatste zinsplaats	28.82% (438/1520)	71.18% (1082/1520)
Met laatste zinsplaats	40.34% (351/870)	59.66% (519/870)

In eerste instantie valt op dat meer dan een derde van alle zinnen in de dataset een laatste zinsplaats bevatten (870/2390, 36.40%). Dat aantal komt grosso modo overeen met wat bekend is uit vroegere corpusstudies (al is daar vaak ander materiaal gebruikt en komt de onderzoekscontext niet altijd overeen met wat hier als onderzoekscontext gedefinieerd is): zo bevat 28.71% (n = 404) van de gevallen in de dataset van De Schutter (1996) een laatste zinsplaats en 20.15% van de gevallen in De Schutter (1976) (n = 814; enkel De Schutters groepen 1 en 2 zijn hier in de vergelijking betrokken).

Belangrijker is echter dat de aanwezigheid van een laatste zinsplaats de keuze voor de groene volgorde [PART+AUX] inderdaad lijkt te stimuleren, terwijl de afwezigheid van een laatste zinsplaats het gebruik van de rode volgorde [AUX+PART] bevordert. In vergelijking met het globale percentage groene [PART+AUX] en rode woordvolgordes [AUX+PART], respectievelijk 33.01% en 66.99%, is dat een verschil van 4.19% voor de gevallen zonder een laatste zinsplaats en 7.39% voor de gevallen met een laatste zinsplaats. Het geobserveerde verschil tussen aan- en afwezigheid van een laatste zinsplaats is statistisch significant ($\chi^2 = 33.26$, $df = 1$, $p < .0001$), hetgeen betekent dat we voor meer dan 99% zeker kunnen zijn dat er een echt verband is tussen de keuze van de woordvolgorde en het al dan niet voorkomen van een laatste zinsplaats. De sterkte van de associatie, uitgedrukt in termen van een o.r., is 0.6 (B.I. = 0.5-0.72); dat betekent dat de kans op het voorkomen van een groene volgorde [PART+AUX] (versus een rode woordvolgorde [AUX+PART]) ($1 / 0.6 =$) 1.66 keer verkleint wanneer de verbale cluster de zin afsluit versus wanneer een laatste zinsplaats volgt op de verbale cluster.

Die resultaten bevestigen de bevindingen van De Schutter (1976, 1996; [HYP 4.9]) en lijken er met andere woorden op te wijzen dat ook in de rechtercontext een accentueringsprincipe werkzaam is die de keuze van de woordvolgorde in de eindgroep beïnvloedt: wanneer een laatste zinsplaats, waarvan verondersteld wordt dat ze gefocaliseerd is en dus een zwaar accent draagt, volgt op de werkwoordelijke eindgroep, dan stijgt de voorkeur voor de groene volgorde [PART+AUX], omdat zo een accentenbotsing vermeden kan worden (het onbeklemtoonde hulpwerkwoord komt tussen het deelwoordaccent en het accent van de laatste zinsplaats).

Hierboven hebben we er al enkele keren op gewezen dat een aantal aspecten van het accentueringsprincipe veeleer vaag blijven en daarom verdere uitdieping behoeven. Naast de algemene vragen en opmerkingen over de aard en de identificatie van zinsaccenten en de toepasbaarheid op geschreven taal (zie paragraaf 4.1), kan men zich in deze context ook nog afvragen of alle laatste zinsplaatsen sowieso in focus staan en, indien niet, hoe focale van niet-focale laatste zinsplaatsen onderscheiden kunnen worden.³¹ Daarenboven kan men zich ook vragen stellen bij de assumptie dat de volledige laatste zinsplaats als accentdragende eenheid beschouwd kan worden. Een dergelijke verificatie van het accentueringsprincipe kan o.i. niet volgehouden

worden indien het accentueringsprincipe georganiseerd is in functie van de notie *accent*, dat “associates to a stress-prominent *syllable* in a word” (Selkirk 1995: 551; onze nadruk) en niet in functie van een pragmatische notie zoals *focus* (Nootboom 1997: 670), waarbij een zinsaccent enkel gebruikt wordt als middel om de gefocaliseerde constituent te detecteren. Uit de definitie van het accentueringsprincipe in paragraaf 4.1 is echter gebleken dat de notie *accent* niet als een middel gebruikt is, maar als centraal element in de definitie (en het is de notie *focus* die gebruikt wordt om het zinsaccent te detecteren).³² Om tegemoet te komen aan die bedenkingen zullen we in paragraaf 4.4.3.3 een meer gerichte prosodische analyse uitvoeren, maar eerst zal een verfijnde syntactische operationalisering van de invloed van de accentuering in de rechtercontext onder de loep genomen worden, met name de invloed van de grammaticale relatie.

4.4.3.2 Grammaticale relatie tussen de laatste zinsplaats en het hoofd

In een tweede stap willen we nagaan of de woordvolgorde in de tweeledige werkwoordelijke eindgroep beïnvloed wordt door de grammaticale relatie tussen de laatste zinsplaats en het hoofd. We onderscheiden daarbij, gebaseerd op wat gebruikelijk is in syntactisch onderzoek, tussen adjuncten en complementen.³³ Alhoewel het onderscheid tussen complementen en adjuncten op het eerste gezicht niets met accentuering te maken lijkt te hebben, hebben onderzoekers, waaronder Braecke (1990: 129-130), de hypothese geformuleerd dat complementen, in tegenstelling tot adjuncten, meer geschikt zijn om in focus te staan en daardoor vaker beklemtoond worden. De hypothese is dan ook de volgende:

[HYP 4.10] Een postverbaal complement (versus een postverbaal adjunct) doet de voorkeur voor de groene volgorde [PART+AUX] stijgen.

Vooraleer over te gaan tot de bespreking van de resultaten, staan we eerst nog even stil bij de wijze waarop we het onderscheid tussen complementen en adjuncten geoperationaliseerd hebben.

Operationalisering van de factor

Laten we beginnen met op te merken dat een complement en een adjunct met elkaar gemeen hebben dat ze een begeleidende functie hebben ten aanzien van het hoofd of de kern van een zin of constituent. Een complement en een adjunct verschillen dan weer van elkaar doordat de eerste groep het hoofd semantisch completeert en door het hoofd geselecteerd wordt, terwijl de tweede groep semantisch en syntactisch een losse toevoeging is bij het hoofd. Tot de complementen worden bijvoorbeeld het subject en de objecten van de zin gerekend, omdat het aantal en de aard van deze elementen (subject en objecten) afhangt van het hoofdwerkwoord dat de complementen selecteert. Typische adjuncten zijn bijwoordelijke bepalingen. In de voorbeeldzinnen (21) en (22) functioneert de laatste zinsplaats respectievelijk als complement (direct object) en adjunct (bijwoordelijke bepaling van plaats):

- (21) [...] dat men te goeder trouw heeft aangenomen dat het om een gelijkwaardig universitair diploma gaat. (DS)
- (22) [...] dat de gesmokkelde diamanten gestolen waren uit de diamantmijnen in Kimberley en aan de Westkust. (DS)

Om de vraag naar de grammaticale relatie tussen de laatste zinsplaats en het hoofd te kunnen beantwoorden, moeten we dus eerst uitmaken wat het hoofd is van de laatste zinsplaats. Anders dan in het voorbeeld dat we net gaven, hoeft een laatste zinsplaats niet geselecteerd te worden door het hoofdwerkwoord, i.c. het deelwoord in de tweeledige eindgroep (zoals in voorbeeldzin (21)), maar kan het ook geselecteerd worden door een preverbaal nomen, adjectief of adverbium:

- (23) [...] dat precies hier het verschroeide [lijk]_{hoofd} werd gevonden van een Afrikaan die al dood geweest moet zijn toen het vuur uitbrak? (DS)

Om uit te maken wat het hoofd van de laatste zinsplaats is, werd de volgende test ontwikkeld:

Transformeer de zin, met weglating van de laatste zinsplaats, in een wat/welk(e)/hoe-vraagzin. Indien het antwoord op die vraag kan bestaan uit de

laatste preverbale constituent en de postverbale constituent, zonder dat dat nonsensicaliteit of ongrammaticaliteit tot gevolg heeft, dan is het hoofd van de laatste zinsplaats preverbaal gesitueerd. Heeft de transformatie ongrammaticaliteit of nonsensicaliteit tot gevolg, dan is het hoofd van de zin het hoofdwerkwoord (i.c. het deelwoord in de eindgroep).

Toegepast op zin (21) en (23) geeft dat het volgende resultaat (* staat voor ‘nonsensicaal’ of ‘ongrammaticaal’):

- (21’) Wat heeft men te goeder trouw aangenomen? *Te goeder trouw dat het om een gelijkwaardig universitair diploma gaat. => [hoofd = V]
 (23’) Welk verschroeid lijk werd precies hier gevonden? Het verschroeiende lijk van een Afrikaan die al dood geweest moet zijn toen het vuur uitbrak. [hoofd = N]

Als eenmaal het hoofd van de laatste zinsplaats geïdentificeerd is, kan de aard van de relatie tot het hoofd bepaald worden. Om uit te maken of een laatste zinsplaats al dan niet geselecteerd wordt door het hoofd, hebben verschillende taalwetenschappers in diverse theoretische frameworks (o.a. Jackendoff 1977, Pollard & Sag 1987, Helbig & Schenkel 1991) criteria bedacht. Niet alle criteria uit de linguïstische literatuur zijn echter te allen tijde bruikbaar in een studie die gebaseerd is op niet-geëliciteerd en niet-introspectief verkregen taalmateriaal. Zelfs recente corpuslinguïstische ontwikkelingen, die een gunstige invloed gehad hebben op de praktische kwaliteit van de tests, resulterend in parsers die automatisch grammaticale relaties kunnen herkennen (Briscoe & Carroll 1997, Buchholz 2002) en grammaticale woordenboeken (Comlex: Grishmann, Macleod & Meyers 1994), hebben niet geleid tot een batterij criteria die de grammaticale relatie tussen eender welke eenheden in de natuurlijke taal eenduidig kan bepalen. De belangrijkste struikelblokken zijn de volgende: (i) verschillende tests leveren tegengestelde resultaten op (zie bv. Jacobs 1994), omdat ze vaag gedefinieerd zijn (met multi-interpretabiliteit als gevolg) en daarom heel erg afhangen van de grammaticaliteitsoordelen van de analist (bv. bij de toepassing van het concept ‘natuurlijkheid’ in de *fronting test*; Meyers, Macleod & Grishman 1996). (ii) De tests zijn meestal context-onafhankelijk opgesteld, zodat de toepassing ervan op reëel geattesteerd taalmateriaal niet altijd even makkelijk is (bv.: kan een thematische rol als ‘instrument’ altijd als een complement gecategoriseerd worden of

zijn er (contextuele) beperkingen?) of zelfs gewoon ontoepasbaar wordt (bv. de test van de geïmpliceerde betekenis; Meyers, Macleod & Grishman 1996). (iii) De meeste tests zijn gericht op het Engels of Duits, waardoor de tests soms moeilijk toe te passen zijn op Nederlandstalige zinnen. Tests om voorlopige en eigenlijke onderwerpen te identificeren, bijvoorbeeld, zijn niet voorhanden. Toch hebben we getracht om op basis van de verwezenlijkingen op corpuslinguïstisch gebied en de beschrijving in ANS (1997: 1366-1385, 845-874, 892-904, 908-914) een theorie-onafhankelijke set van criteria samen te stellen die elke relatie tussen de laatste zinsplaats en het hoofd moet kunnen benoemen.³⁴ In appendix C worden die criteria opgesomd.

Resultaten

In tabel 4.16 wordt het resultaat van de toepassing van die criteria op de zinnen in onze dataset die een laatste zinsplaats bevatten (n = 870 of 36.40% van het totaal) samengevat:

Tabel 4.16: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per grammaticale relatie tussen de laatste zinsplaats en het hoofd.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Complement	47.54% (251/528)	52.46% (277/528)
Adjunct	29.24% (100/342)	70.76% (242/342)

Wat in de eerste plaats uit tabel 4.16 blijkt, is dat van de 870 zinnen met een laatste zinsplaats er 528 (60.69%) een complement (met een werkwoord of een preverbale constituent als hoofd) als laatste zinsplaats hebben en 342 (39.31%) een adjunct als laatste zinsplaats. Dat betekent met andere woorden dat taalgebruikers vaker een constituent extraponeren die door het hoofd geselecteerd wordt dan een constituent die als losse toevoeging aan de zin is gehecht. Het proportionele verschil tussen complementen en adjuncten is statistisch significant ($\chi^2 = 39.34$, $df = 1$, $p < .0001$), wat betekent dat met 99.99% zekerheid aangenomen kan worden dat complementen vaker geëxtraponeerd worden dan adjuncten.³⁵

Naast het opvallende verschil in extrapositiegedrag tussen complementen en adjuncten laat tabel 4.16 in de tweede plaats ook zien dat de voorkeur voor een van de woordvolgordevarianten lijkt af te hangen van de grammaticale relatie tussen de laatste zinsplaats en het hoofd, en wel in die zin dat de keuze voor de groene volgorde [PART+AUX] gestimuleerd wordt wanneer de werkwoordelijke eindgroep gevolgd wordt door een complement (47.54%)—in vergelijking met een adjunct (29.24%). Een chi-kwadraatanalyse laat zien dat de associatie tussen grammaticale relatie en woordvolgorde significant is ($\chi^2 = 28.87$, $df = 1$, $p < .0001$). De sterkte van de associatie, uitgedrukt in o.r.'s, geeft te kennen dat de waarschijnlijkheid voor de groene woordvolgorde [PART+AUX] versus de rode volgorde [AUX+PART] 2.19 keer hoger ligt wanneer de verbale cluster gevolgd wordt door een complement dan wanneer de eindgroep gevolgd wordt door een adjunct (B.I. = 1.63-2.96). Aangenomen dat complementen inderdaad vaker in focus staan en daardoor vaker beklemtoond worden (Braecke 1990: 129-130), bevestigen de resultaten opnieuw het bestaan van een accentueringsprincipe in de rechtercontext: om een accentenbotsing tussen het beklemtoonde deelwoord en het beklemtoonde postverbale complement te vermijden, wordt daarom de groene volgorde [PART+AUX] gekozen, met het onbeklemtoonde hulpwerkwoord tussen het deelwoord en het complement in (zie echter de opmerkingen die in paragraaf 4.4.3.1 gemaakt zijn).

In de rest van deze paragraaf gaan we nog na of het type hoofd dat de laatste zinsplaats selecteert ook een effect heeft op de woordvolgorde, i.e. of het verschil tussen een werkwoordelijk hoofd (i.c. het deelwoord) en een preverbaal niet-werkwoordelijk hoofd een invloed heeft op de keuze voor de groene [PART+AUX] of de rode [AUX+PART] woordvolgorde:

(24) [...] dat men te goeder trouw heeft [aangenomen]_{hoofd} [dat het om een gelijkwaardig universitair diploma gaat]. (DS)

(25) [...] dat precies hier het verschroeide [lijk]_{hoofd} werd gevonden [van een Afrikaan die al dood geweest moet zijn toen het vuur uitbrak]? (DS)

In tabel 4.17 worden daartoe de gegevens voor de beide types hoofden verdeeld over de woordvolgordevarianten:

Tabel 4.17: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type hoofd van de laatste zinsplaats.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
preverbaal hoofd	17.22% (26/151)	82.78% (125/151)
verbaal hoofd	45.20% (325/719)	54.80% (394/719)

Ook al is de rode volgorde [AUX+PART] de dominante woordvolgordevariant, toch kan niet voorbijgegaan worden aan het immense verschil tussen de beide types hoofden: zo blijkt uit tabel 4.17 dat een laatste zinsplaats met een preverbaal hoofd de keuze voor de rode volgorde [AUX+PART] gigantisch doet stijgen in vergelijking met een laatste zinsplaats die tot het valentieschema van het hoofdwerkwoord behoort. Het verschil tussen de beide types is erg significant ($\chi^2 = 40.60$, $df = 1$, $p < .0001$) en de sterkte van de associatie tussen het type hoofd en de woordvolgorde, uitgedrukt in o.r., is 0.25 (B.I. = 0.15-0.40), hetgeen betekent dat de kans op een groene volgorde [PART+AUX] (versus de kans op een rode volgorde [AUX+PART]) ongeveer 4 keer kleiner is als de laatste zinsplaats een preverbaal hoofd heeft, vergeleken met een laatste zinsplaats met een verbaal hoofd.

Aan de hand van de resultaten kan nu in de eerste plaats geconcludeerd worden dat, net zoals bij de analyse in termen van adjuncten en complementen, de categorie *laatste zinsplaats* in het kader van het onderzoek naar de woordvolgordevariatie in de werkwoordelijke eindgroep onder geen beding opgevat mag worden als een homogene categorie. In de tweede plaats kan aan de hand van de resultaten in tabel 4.17 afgeleid worden dat de syntactische aanhechtingssite van de laatste zinsplaats van belang blijkt te zijn bij de bepaling van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde of, anders geformuleerd, dat taalgebruikers bij de keuze voor de rode [AUX+PART] of de groene volgorde [PART+AUX] zich op zijn minst gedeeltelijk laten leiden door de aard (werkwoordelijk of niet-werkwoordelijk hoofd) en/of de positie van het hoofd van de laatste zinsplaats. Bij een verbaal hoofd heeft de laatste zinsplaats nl. onmiddellijk aansluiting op het hoofd—een eventueel hulpwerkwoord in

de groene woordvolgorde [PART+AUX] uitgezonderd—terwijl een preverbaal hoofd altijd van de laatste zinsplaats gescheiden wordt door de werkwoordelijke eindgroep:³⁶

(26) [...] dat men te goeder trouw heeft [aangenomen]_{hoofd} [dat het om een gelijkwaardig universitair diploma gaat]. (DS)

(27) [...] dat precies hier het verschroeide [lijk]_{hoofd} werd gevonden [van een Afrikaan die al dood geweest moet zijn toen het vuur uitbrak]? (DS)

In tegenstelling tot de interpretatie van het effect van de grammaticale relatie in termen van adjuncten en complementen, ligt een interpretatie van het effect van grammaticale relatie in termen van types hoofden veel minder voor de hand: volgens Shannon (1995) staat een laatste zinsplaats met een preverbaal hoofd nl. in focus, waardoor men aan de hand van het accentueringsprincipe zou verwachten dat de groene volgorde [PART+AUX] in die omstandigheden de voorkeur wegdraagt, omdat op die manier het onbeklemtoonde hulpwerkwoord tussen het beklemtoonde deelwoord en de beklemtoonde laatste zinsplaats geplaatst wordt en een accentenbotsing vermeden kan worden. De resultaten laten echter net het omgekeerde beeld zien, m.n. een erg sterke voorkeur voor de rode volgorde [AUX+PART], zodat het effect van het hoofd van de laatste zinsplaats, in tegenstelling tot het effect van de aard van grammaticale relatie, onmogelijk verklaard lijkt te kunnen worden in het kader van het accentueringsprincipe. In hoofdstuk 7 zullen we een poging ondernemen om het effect op een andere manier te interpreteren.

Vergelijking van de analyses

Tot slot van deze paragraaf kunnen we ons nog de vraag stellen naar de precieze verhouding van de variabelen *grammaticale relatie* en *type hoofd*. Totnogtoe zijn we ervan uitgegaan dat die beide variabelen onafhankelijk van elkaar een effect hebben op de woordvolgorde, maar het is uiteraard niet uitgesloten dat ze allebei net hetzelfde effect hebben. Dat zou bijvoorbeeld het geval zijn als zou blijken dat een laatste zinsplaats die afhankelijk is van een preverbaal hoofd vaker als adjunct gebruikt wordt (voorbeeldzin (28)) en dat, mutatis mutandis, een laatste zinsplaats die afhankelijk is van een verbaal hoofd vaker als complement verschijnt (voorbeeldzin (29)):

- (28) [...] dat de Franse luchtvaartmaatschappij K's Choice op het rockbandje heeft gezet dat men tijdens lange vluchten kan beluisteren. (DS)
- (29) [...] dat geen aandacht werd besteed aan de financiering van het plan. (DS)

Om de relatie tussen de grammaticale relatie en het type hoofd in hun effect op de woordvolgorde na te gaan, zullen we daarom opnieuw een conditionele analyse uitvoeren: in een eerste stap zal het effect van de grammaticale relatie op de woordvolgorde herbestudeerd worden onder controle van het type hoofd, in een tweede stap worden de rollen omgedraaid en wordt met andere woorden het effect van het type hoofd op de woordvolgorde geheranalyseerd onder controle van de grammaticale relatie.

De eerste CMH-analyse maakt duidelijk dat de globale associatie (i.e. een samenvatting van de partiële analyses) tussen het type grammaticale relatie en de woordvolgorde, gecontroleerd voor het type hoofd, statistisch significant is (CMH- $\chi^2 = 39.98$, $df = 1$, $p < .0001$). De globale o.r. is 0.39 (B.I. = 0.29-0.53; Br-D $\chi^2 = 0.64$, $df = 1$, $p > .05$). Dat betekent dat, globaal genomen, het onderscheid tussen complementen en adjuncten relevant is. Bij een nadere bestudering van de partiële analyses blijkt wel dat er een interactie-effect is tussen het effect van de grammaticale relatie en de laatste zinsplaatsen die een verbaal hoofd hebben ($\chi^2 = 38.89$, $df = 1$, $p < .0001$; o.r. = .38, B.I. = 0.28-0.51). Dat betekent dat het onderscheid tussen adjuncten en complementen niet belangrijk is in het geval een laatste zinsplaats een preverbaal hoofd heeft; dergelijke types bevorderen de voorkeur voor de rode volgorde [AUX+PART] altijd erg sterk. Een linguïstische interpretatie van het interactie-effect ligt niet meteen voor de hand, maar wat wel duidelijk is, is dat het interactie-effect te wijten zou kunnen zijn aan de lage absolute frequenties van de categorie *preverbaal hoofd* (waardoor het mogelijk is dat we te weinig statistische kracht hebben om een effect in alle partiële tabellen te verkrijgen). In toekomstig onderzoek zal dat zeker verder onderzocht moeten worden.

Als we de rollen van de verklarende en controlevariabele omdraaien, dan zien we opnieuw een globaal significant effect opduiken (CMH- $\chi^2 = 50.10$, $df = 1$, $p < .0001$). Een schatting van de globale o.r. is 0.21 (B.I. = 0.14-0.33), de Breslow-Day test wijst op homogene o.r.'s in de partiële analyses (Br-D $\chi^2 = 0.64$, $df = 1$, $p > .05$),

waardoor de globale o.r. een goede samenvatting geeft van de partiële o.r.. In tegenstelling tot de vorige CMH-analyse zijn er geen interactie-effecten en is, na bestudering van de partiële analyses, het onderscheid tussen preverbaal en verbaal hoofd een relevant en statistisch significant onderscheid (gecontroleerd voor adjuncten: $\chi^2 = 6.54$, $df = 1$, $p = .01$; o.r. = .30, B.I. = 0.11-0.79; gecontroleerd voor complement: $\chi^2 = 45.09$, $df = 1$, $p < .0001$; o.r. = .19, B.I. = 0.12-0.32). Daaruit blijkt dus dat de beide variabelen relevant zijn in het onderzoek en dat ze de keuze voor de groene [PART+AUX] of rode [AUX+PART] volgorde elk op hun eigen, unieke manier beïnvloeden.

4.4.3.3 *Afstand van het deelwoordaccent tot het eerste postverbale woordaccent*

In een laatste, specifiek prosodische analyse van de impact van de accentverdeling in de rechtercontext werd het aantal syllabes tussen het woordaccent van het deelwoord en het woordaccent van het eerste postverbale woord geteld (de zinnen zonder een laatste zinsplaats worden buiten beschouwing gelaten). De identificatie van de woordaccenten gebeurde aan de hand van de principes die in paragraaf 4.3.3.5 voorgesteld zijn. Het aantal syllabes van het hulpwerkwoord, die kunnen interveniëren tussen het deelwoord en het eerste postverbale woord in het geval van de groene woordvolgorde [PART+AUX], werd niet meegeteld, om een incorporatie van de responsvariabele in de verklarende variabele te vermijden (zie paragraaf 4.2.3.3 en 4.3.3.5). De hypothese is de volgende:

[HYP 4.11] Naarmate de afstand tussen het deelwoordaccent en het eerste postverbale accent kleiner wordt, worden meer groene woordvolgordes [PART+AUX] geprefereerd.

De redenering achter [HYP 4.11] is dat groene woordvolgordeclusters [PART+AUX], met het onbeklemtoonde hulpwerkwoord tussen het postverbale accent en het deelwoordaccent, een accentenbotsing kunnen vermijden. Tabel 4.18 vat het effect van de woordaccentdistributie in de rechtercontext op de woordvolgorde samen:

Tabel 4.18: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per aantal tussenliggende ongeaccentueerde lettergrepen tussen het deelwoordaccent en het eerste postverbale woordaccent.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
0 ongeaccentueerde syllabes	51.01% (202/396)	48.99% (194/396)
1 ongeaccentueerde syllabes	38.96% (97/249)	61.04% (152/249)
2 ongeaccentueerde syllabes	25.44% (29/114)	74.56% (85/114)
3-4 ongeaccentueerde syllabes	20.72% (23/111)	79.28% (88/111)

De resultaten in tabel 4.18 geven aan, zoals voorspeld door het accentueringsprincipe, dat er een positieve lineaire trend is, nl. hoe groter de afstand tussen de geaccentueerde syllabes van het deelwoord en de eerste geaccentueerde syllabe op het postverbale woord, hoe meer rode woordvolgordes [AUX+PART]. Dat betekent met andere woorden dat in het geval van een klein interval tussen de twee geaccentueerde syllabes de groene woordvolgorde [PART+AUX] frequenter gekozen wordt, in het geval van een groter interval wordt de rode volgorde [AUX+PART] vaker gekozen. Omdat hier sprake is van een ordinale verklarende variabele zullen we opnieuw gebruik maken van de Mantel-Haenszel chi-kwadraattest om na te gaan of de lineaire trend statistisch significant is. En dat blijkt inderdaad zo te zijn ($MH-\chi^2 = 46.17$, $df = 1$, $p < .0001$); de γ -coëfficiënt, zoals bekend een maat voor de sterkte van de ordinale associatie, is gelijk aan 0.37 (ASE = 0.05): hoe groter het interval, hoe meer rode woordvolgordes [AUX+PART] (bevestiging [HYP 4.11]).

Ondanks het feit dat de operationalisering van de variabele niet optimaal is—we hebben de afstand tussen woordaccenten en niet tussen zinsaccenten geteld—waren we toch in staat om het bestaan van het accentueringsprincipe te verifiëren. Toch kan men tegen deze operationalisering nog inbrengen dat het niet ideaal was om kritiekloos het woordaccent van het eerste postverbale woord te selecteren, omdat dat heel vaak een functiewoord is (slechts 59/870, 6.78% van alle observaties met een

laatste zinsplaats hadden een inhoudswoord als eerste postverbale woord!), zoals *voor* in voorbeeldzin (30):

(30) [...] dat de onderzeeër *ingezet is voor* een infiltratieopdracht. (DS)

Het is hoogst twijfelachtig of woordaccenten van functiewoorden zo prominent zijn als woordaccenten van inhoudswoorden, en daarom zou men zich kunnen afvragen of de afstand tussen het accent op het deelwoord en het eerste postverbale woord geteld moet worden, dan wel de afstand tussen deelwoord en eerste postverbale inhoudswoord. Om aan die kritiek tegemoet te komen, wordt een bijkomende analyse uitgevoerd waarbij het aantal lettergrepen tussen het woordaccent van het deelwoord en het woordaccent van het eerste postverbale inhoudswoord (i.e. een werkwoord, nomen, pronomen, adjectief, adverbium of telwoord) geteld werd (opnieuw werd het hulpwerkwoord genegeerd in de telling).

Tabel 4.19: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per aantal tussenliggende ongeaccentueerde lettergrepen tussen het deelwoordaccent en het accent van het eerste postverbale inhoudswoord.

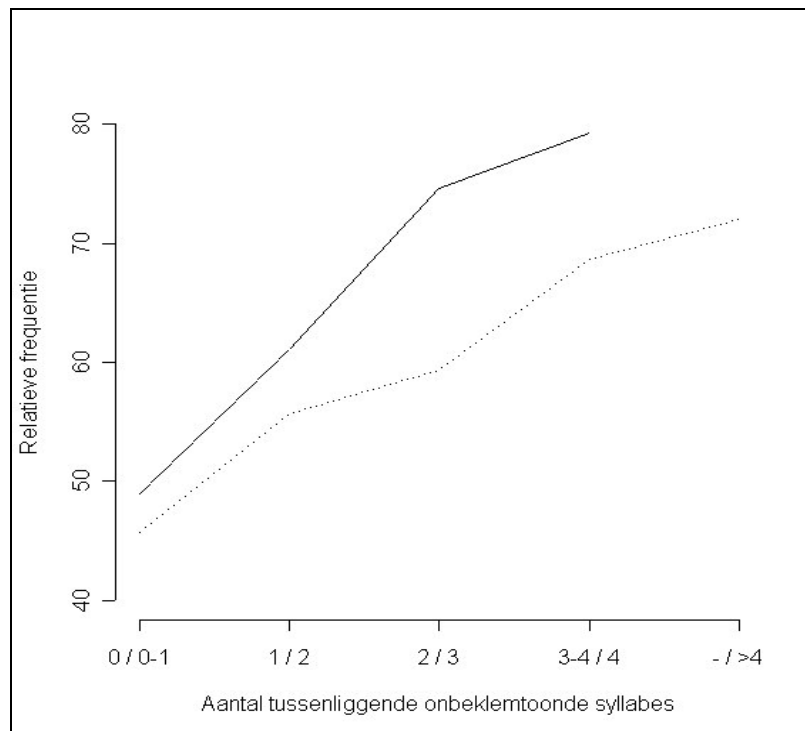
	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
0-1 ongeaccentueerde syllabes	54.30% (81/151)	45.70% (70/151)
2 ongeaccentueerde syllabes	44.35% (102/230)	55.65% (128/230)
3 ongeaccentueerde syllabes	40.69% (84/205)	59.31% (121/205)
4 ongeaccentueerde syllabes	31.34% (42/134)	68.65% (92/134)
>4 ongeaccentueerde syllabes	28.00% (42/150)	72.00% (108/150)

Dezelfde lineaire trend als in tabel 4.18 duikt ook hier op: de groene volgorde [PART+AUX] wordt geprefereerd als er weinig onbeklemtoonde ruimte is tussen twee zwaar geaccentueerde syllabes (54.30%), de rode volgorde [AUX+PART] wordt

gestaag populairder naarmate de onbeklemtoonde ruimte groter wordt (72.00%). De lineaire trend is statistisch significant ($MH-\chi^2 = 25.84$, $df = 1$, $p < .0001$; $\gamma = 0.25$, $ASE = 0.05$). Het accentueringsprincipe kan hier dus opnieuw bevestigd worden als verklarend mechanisme voor de woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen (bevestiging [HYP 4.11]).

Figuur 4.4 vat de resultaten van de positieve lineaire trend van de beide operationalisering van de accentdistributie nog eens visueel samen (de gestippelde lijn geeft de resultaten van de eerste operationalisering weer, de volle lijn representeert de resultaten van de tweede operationalisering):

Figuur 4.4: Relatieve proporties van de rode [AUX+PART] woordvolgorde per afstand van het deelwoordaccent tot het woordaccent van het eerst postverbale (inhouds)woord (in termen van tussenliggende onbeklemtoonde syllabes).



Ondanks die duidelijke trend moet toch opgemerkt worden dat het accentueringsprincipe onmogelijk een verklarend mechanisme kan zijn dat absoluut dwingend is, ofwel moet de formulering van het accentueringsprincipe aangepast worden. Uit de resultaten van ons onderzoek is nl. gebleken dat, in tegenstelling tot wat verwacht werd, ongeveer de helft (48.99% en 45.70% voor respectievelijk tabel 4.18 en 4.19) van de observaties die weinig onbeklemtoonde syllabes (0 en 0/1) tussen

de twee woordaccenten hebben in de rode woordvolgorde [AUX+PART] verschijnt. Omgekeerd blijkt het in de praktijk van het taalgebruik hoegenaamd niet minder ideaal te zijn (De Schutter 1996: 215) om ruime intervallen (van meer dan 3 onbeklemtoonde syllabes) tussen de twee woordaccenten te hebben: maar liefst 20.72% en 59.34% van de observaties met een ruim interval duikt niet in de rode woordvolgorde [AUX+PART] op. Daarenboven moet ook nog in rekening worden gebracht dat onze operationalisering gebaseerd zijn op woordaccenten en niet op zinsaccenten. Als men aanneemt dat een zin meer woordaccenten dan zinsaccenten bevat, dan zou een operationalisering in termen van zinsaccenten wellicht nog grotere onbeklemtoonde intervallen opleveren.

4.4.3.4 *Vergelijking van de verschillende operationalisering*

Het onderzoek naar de invloed van de accentuering in de rechtercontext op de woordvolgorde in de werkwoordelijke eindgroep heeft in de eerste plaats duidelijk gemaakt dat er een effect is van de aan- of afwezigheid van de laatste zinsplaats, in die zin dat de voorkeur voor de groene volgorde [PART+AUX] groter is als er een laatste zinsplaats volgt. Dat resultaat ligt volledig in de lijn van het accentueringsprincipe, omdat het onbeklemtoonde hulpwerkwoord tussen het beklemtoonde deelwoord en de beklemtoonde laatste zinsplaats in komt te staan en op die manier een accentenbotsing vermijdt. Daarnaast werd de invloed van de accentdistributie ook onderzocht door de categorie *laatste zinsplaats* te differentiëren naar grammaticale relatie. De resultaten van een van de analyses die we in dat kader uitgevoerd hebben, bevestigden eveneens het bestaan van een accentueringsprincipe: postverbale complementen, waarvan aangenomen wordt dat ze in focus staan en daarom een accent dragen, bleken nl. vaker voorafgegaan te worden door een groene werkwoordsvolgorde [PART+AUX] dan postverbale adjuncten, die geen accent dragen (de groene volgorde [PART+AUX] vermijdt een accentenbotsing tussen het beklemtoonde deelwoord en het beklemtoonde postverbale complement). Tot slot hebben we de accentdistributie ook op een specifiekere manier onderzocht, met name door de afstand tussen de woordaccenten van het deelwoord en het eerste postverbale (inhouds)woord te tellen. De resultaten van die analyse bevestigden eveneens het bestaan van een accentueringsprincipe in de keuze voor een van de woordvolgordes: de keuze voor de groene volgorde [PART+AUX] stijgt naarmate er minder onbeklemtoonde syllabes

tussen het woordaccent van het deelwoord en dat van het eerste postverbale (inhouds)woord zitten.

Alle operationaliseringën, met uitzondering van die in termen van types hoofden, lijken met andere woorden in staat om het bestaan van een accentueringsprincipe aan te tonen. De vraag die men zich daarbij kan stellen, is uiteraard in hoeverre alle operationaliseringën dat in gelijke mate doen. Het hoeft nauwelijks betoog dat de operationalisering in termen van tussenliggende onbeklemtoonde lettergrepen in staat is om het principe veel specifieker te testen en in meer van zijn aspecten dan de operationalisering in termen van al dan niet aanwezigheid van een laatste zinsplaats, ook al is de meest specifieke analyse die hier uitgevoerd is nog niet specifiek genoeg om het principe te verifiëren zoals het oorspronkelijk bedoeld is. De operationalisering in termen van een syntactico-pragmatische categorie als laatste zinsplaats of postverbaal complement/adjunct is zo ruw—men neemt aan dat volledige constituenten, zelfs volledige bijzinnen accentdragende eenheden zijn—dat de verschillende categorieën uit heel heterogene elementen bestaan en onderliggende mechanismen mogelijk het globale beeld vertekenen.

Een nadere beschouwing van de resultaten van de operationalisering in termen van aan- vs. afwezigheid van een laatste zinsplaats en de operationalisering in termen van tussenliggende onbeklemtoonde syllabes in tabel 4.15 en 4.18/4.19 moge duidelijk maken waar we op doelen. In tabel 4.15 werden de zinnen met en de zinnen zonder een laatste zinsplaats tegen elkaar afgezet, in tabel 4.18 en 4.19 werd enkel gekeken naar het gedeelte van de dataset met zinnen die een laatste zinsplaats hebben (die zonder een laatste zinsplaats werden buiten beschouwing gelaten). Het inzicht dat we verkregen in het accentueringsprincipe na bestudering van tabel 4.18 en 4.19 hadden we met andere woorden niet verkregen in tabel 4.15, omdat geen verdere onderscheidingen aangebracht waren. Wat echter belangrijker is, is niet zozeer dat in tabel 4.18 en 4.19 extra onderscheidingen aangebracht zijn, maar wel dat die onderscheidingen duidelijk gemaakt hebben dat binnen een syntactisch afgebakende heterogene categorie als *laatste zinsplaats* verschillende tendensen aanwezig kunnen zijn: een zin met een laatste zinsplaats waarbij het eerste woordaccent van het eerste postverbale (inhouds)woord zich op korte afstand van het woordaccent van het

voorafgaande deelwoord bevond, had een grotere voorkeur voor de groene volgorde [PART+AUX], dan een zin met een laatste zinsplaats waarbij het eerste woordaccent van het eerste postverbale (inhouds)woord zich op langere afstand van het woordaccent van het deelwoord bevond. Kortom, een globale categorie als *laatste zinsplaats* bleek bij nader inzien verschillende subtypes te bevatten die in de huidige omstandigheden hoogst relevant zijn, zodat men zich de vraag kan stellen of een operationalisering op basis van syntactische gronden überhaupt als een operationalisering van een prosodisch principe beschouwd kan worden. Dat hoeft niet te impliceren dat we het effect van die onspecifieke variabele moeten negeren, het betekent enkel dat we het geobserveerde effect misschien op een andere wijze moeten interpreteren dan in termen van prosodie (zie hoofdstuk 7 voor een poging daartoe).

Ook de beide operationalisering in termen van grammaticale relatie lijden onder dezelfde beperkingen als de interpretatie van de aan- versus afwezigheid van een laatste zinsplaats, aangezien opnieuw een syntactisch afgebakende categorie, onafhankelijk van de precieze accentverdeling, gebruikt is om een prosodisch principe te verifiëren. Als men even aanneemt (i) dat een prosodische hypothese prosodisch getest moet worden en (ii) dat de meest geschikte operationalisering—rekening houdend met het type materiaal dat we analyseren—van op zijn minst een van de aspecten van het accentueringsprincipe de woordaccentdistributie is en (iii) dat het effect van de grammaticale relatie en het type hoofd beter niet verklaard wordt in termen van prosodie, maar tegen de achtergrond van een ander, nog te preciseren syntactico-pragmatisch principe en (iv) dat die beide principes naast elkaar kunnen bestaan, dan moet het mogelijk zijn om die coëxistentie met behulp van de gestratificeerde CMH-analyses te verifiëren.

Als er nl. inderdaad twee (of meer) verklarende mechanismen bestaan, dan is de meest voor de hand liggende verwachting dat het streven naar een evenwichtige accentverdeling zowel geldt voor postverbale complementen als voor postverbale adjuncten apart, voor de laatste zinsplaatsen met een verbaal en met een preverbaal hoofd apart en, omgekeerd, moeten de syntactico-pragmatische verschillen waarvan in paragraaf 4.4.3.2 sprake was, ook gelden in contexten met een klein, middelmatig en een groot onbeklemtoond interval. Als het accentueringsprincipe en het syntactico-pragmatische principe net hetzelfde zouden verklaren, doordat bijvoorbeeld het

syntactische principe herleid kan worden tot het accentueringsprincipe, dan kan verwacht worden dat een van de verschillen—het verschil in accentdistributie of het verschil in grammaticale relatie en/of type hoofd—redundant wordt als er gecontroleerd wordt voor het andere verschil.

Een eerste paar conditionele CMH-analyses onderzocht het effect van accentverdeling onder controle van grammaticale relatie en type hoofd. Het tweede paar onderzocht het effect van het type hoofd en de grammaticale relatie onder controle van de accentdistributie.³⁷ In tabel 4.20 wordt een overzicht van de marginale en partiële CMH-statistieken gegeven:

Tabel 4.20: Overzicht van de CMH-resultaten van vier analyses voor het effect van de variabelen *grammaticale relatie*, *type hoofd*, en *accentverdeling*.

Verklarende Variabele	Controle-variabele	Globale CMH-statistieken Globale o.r.	Samenvattende statistieken partiële analyses
Accent	Grammaticale relatie	CMH- $\chi^2 = 22.51$, df = 4, p = .0002	complement: MH $\chi^2 = 16.38$, df = 1, p < .0001; $\gamma = .25$, ASE = 0.06 adjunct: MH $\chi^2 = 5.80$, df = 1, p = .02; $\gamma = .20$, ASE = 0.08
Accent	Type hoofd	CMH- $\chi^2 = 25.37$, df = 4, p < .0001	preverbaal: MH $\chi^2 = 0.04$, df = 1, p > .05 verbaal: MH $\chi^2 = 28.63$, df = 1, p < .0001; $\gamma = .28$, ASE = 0.05
Grammaticale relatie	Accent	CMH- $\chi^2 = 24.76$, df = 1, p < .0001 o.r. = .48, b.i. = .36-.64 Br-D $\chi^2 = .96$, df = 4, p < .05	A: $\chi^2 = 4.05$, df = 1, p = .04; o.r. = .49, B.I. = .24-.99 B: $\chi^2 = 11.09$, df = 1, p = .0009; o.r. = .39, B.I. = .22-.68 C: $\chi^2 = 4.85$, df = 1, p = .03; o.r. = .52, B.I. = .29-.93 D: $\chi^2 = 1.88$, df = 1, p > .05 E: $\chi^2 = 3.92$, df = 1, p = .05; o.r. = .46, B.I. = .22-1
Type hoofd	Accent	CMH- $\chi^2 = 39.37$, df = 1, p < .0001 o.r. = .27, b.i. = .17-.43 Br-D $\chi^2 = 7.51$, df = 4, p > .05	A: $\chi^2 = 10.49$, df = 1, p = .001; o.r. = .18, B.I. = .06-.55 B: $\chi^2 = 18.79$, df = 1, p < .0001; o.r. = .17, B.I. = .07-.41 C: $\chi^2 = 13.54$, df = 1, p = .0002; o.r. = .16, B.I. = .05-.47 D: $\chi^2 = 0.35$, df = 1, p > .05 E: $\chi^2 = 2.39$, df = 1, p > .05

De globale CMH-statistieken geven voor alle analyses een significant effect aan. Dat betekent met andere woorden dat zowel het effect van de accentverdeling als het effect van de grammaticale relatie en het type hoofd, gecontroleerd voor respectievelijk grammaticale relatie/type hoofd en accentverdeling, behouden blijft. Op basis van de samenvattende, globale CMH-statistieken kunnen we dus concluderen dat het accentueringseffect niet zomaar herleid kan worden tot een syntactico-pragmatisch effect en vice versa, waardoor we vastere grond onder de voeten hebben om te beweren dat het erop lijkt dat er minstens twee principes onderscheiden moeten worden in de verklaring van de woordvolgordealternatie in tweeledige werkwoordelijke eindgroepen.

Wanneer we er de resultaten van de partiële analyses bijhalen, wordt het beeld echter complexer, omdat er een aantal interactie-effecten opduikt: zo is het effect van de accentverdeling enkel van toepassing op de laatste zinsplaatsen met een verbaal hoofd en niet op die met een preverbaal hoofd (wat vermoedelijk te wijten is aan dataschaarste). Verder geldt het effect van grammaticale relatie en het effect van het type hoofd niet voor de zinnen waar er een onbeklemtoond interval is van 4 lettergrepen—en voor het effect van het type hoofd ook voor zinnen met een onbeklemtoond interval van meer dan 4 syllaben—tussen het woordaccent van het deelwoord en het woordaccent van het eerste postverbale inhoudswoord.

Wat de reden ook moge zijn voor die interactie-effecten, het is duidelijk dat de prosodische effecten en de syntactico-pragmatische effecten op geen enkele wijze volledig overlappen, zodat we voorlopig kunnen concluderen dat er enerzijds, zoals daarnet al aangekondigd werd, minstens twee verklarende principes lijken te bestaan en anderzijds de effecten van de laatste zinsplaats, de grammaticale relatie en het type hoofd niet zomaar, zoals in vroeger onderzoek gebeurd is, op een prosodische manier geïnterpreteerd mogen worden. Toekomstig onderzoek zal de interactie tussen de beide verder moeten verkennen.

4.5 De interactie tussen de accentverdeling in, voor en na de werkwoordelijke eindgroep

In dit hoofdstuk hebben we een antwoord trachten te bieden op de vraag of er sprake kan zijn van een accentueringsprincipe dat de keuze voor de groene [PART+AUX] of de rode volgorde [AUX+PART] bepaalt. De volgordevariatie in de werkwoordelijke eindgroep wordt daarbij opgevat als een relatief correctiemechanisme dat er enerzijds voor zorgt dat zware accenten niet in elkaars onmiddellijke buurt terechtkomen en anderzijds voldoende prosodische ruimte genereert om het platte-hoedpatroon te realiseren. Daartoe hebben we de accentverdeling in en rond de werkwoordelijke eindgroep bestudeerd en we zijn daar globaal genomen tot de conclusie gekomen dat er inderdaad sprake is van een accentueringsprincipe—al neemt dat niet weg dat nader onderzoek bepaalde aspecten van het principe verder zal moeten verifiëren en eventueel modificeren.

Zo heeft ons onderzoek aangetoond dat naarmate de afstand tussen het deelwoordaccent en een zwaar beklemtoonde lettergreep in het laatste preverbale woord (verder: vorig accent) kleiner wordt, de rode volgorde [AUX+PART] in aanzienlijke mate gestimuleerd wordt; omgekeerd wordt de groene [PART+AUX] volgorde bevorderd naarmate de afstand tussen het deelwoordaccent en het einde van de werkwoordelijke eindgroep of de afstand tussen het deelwoordaccent en het eerste postverbale accent (verder: volgend accent) kleiner is. Die effecten konden gemotiveerd worden vanuit het streven om een accentenbotsing te vermijden. De vraag die we nu nog moeten beantwoorden, is in hoeverre die beide effecten met elkaar in overeenstemming te brengen zijn.³⁸ Er zijn vier basisscenario's:

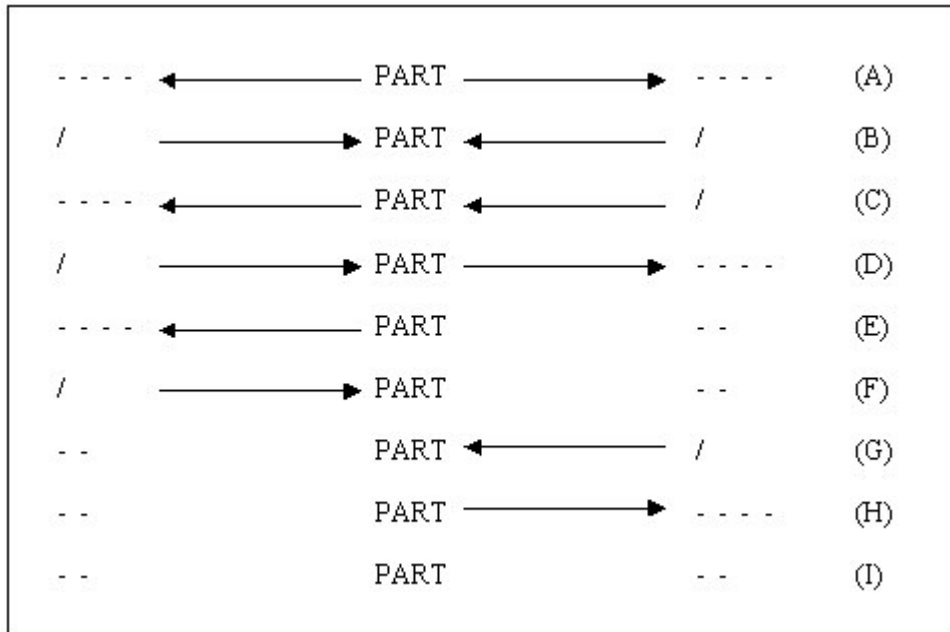
- Er is geen postverbale constituent, waardoor de locatie van het vorige accent bepaalt welke volgorde gekozen wordt, m.n. de rode volgorde [AUX+PART] indien de accenten van het deelwoord en het vorige woord minder dan 2-3 syllabes van elkaar verwijderd zijn, de groene volgorde [PART+AUX] in de andere gevallen.
- De volgorde in de werkwoordelijke eindgroep hangt primair af van de locatie van het accent in het vorige woord, m.n. de rode volgorde [AUX+PART] indien de accenten van het deelwoord en het vorige woord minder dan 2-3 syllabes van elkaar verwijderd zijn, de groene volgorde [PART+AUX] in de

andere gevallen. Dat betekent dat een accentenbotsing tussen het deelwoordaccent en het accent van het volgende woord niet altijd kan worden vermeden.

- De volgorde in de werkwoordelijke eindgroep hangt primair af van de locatie van het volgende accent, m.n. de groene volgorde [PART+AUX] indien de accenten van het deelwoord en het volgende woord minder dan 2-3 syllabes van elkaar verwijderd zijn, de rode volgorde [AUX+PART] in de andere gevallen. Dat betekent dat een accentenbotsing tussen het deelwoordaccent en het accent van het vorige woord niet altijd kan worden vermeden.
- De volgorde in de werkwoordelijke eindgroep hangt zowel af van de locatie van het accent in het vorige woord als van het accent in het volgende woord. Er wordt met andere woorden telkens een compromis gezocht om de accentverdeling voor, in en na de eindgroep zo goed mogelijk op elkaar af te stemmen.

Indien men naast de eis om een accentenbotsing te vermijden (gesymboliseerd in de figuur hieronder door: /³⁹) ook rekening houdt met de eis om geen te lange intervallen met onbeklemtoonde syllabes te hebben (- - -) en met de mogelijkheid dat de afstand tot het vorige of volgende accent ideaal is (- -), dan wordt het beeld uiteraard complexer (de richting van de pijl geeft aan in welke richting het deelwoord geduwd of getrokken wordt):

Figuur 4.5: Negen scenario's voor de interactie tussen het vermijden van een accentenbotsing en een te lang interval met onbeklemtoonde syllabes in en rond de werkwoordelijke eindgroep.



Van de negen mogelijke scenario's, wordt enkel in scenario (C), (D) en (I) een evenwicht tussen de verschillende krachten bereikt. In de andere scenario's is er sprake van een conflict, doordat een te lang interval met onbeklemtoonde lettergrepen niet vermeden kan worden aan een van de uiteinden van de eindgroep (scenario A, E en H) of doordat een accentenbotsing aan een van de uiteinden niet vermeden kan worden (scenario B, F en G). Voor die scenario's kan eveneens de vraag gesteld worden welke kracht primeert.

Aan de hand van een tweetal CMH-analyses zullen we een idee proberen te krijgen hoe die verschillende krachtsverhoudingen op elkaar inwerken. In eerste instantie zullen we daarbij nagaan of het effect van de accentverdeling in de linkercontext even sterk is en in dezelfde richting wijst als de marginale analyse hierboven (figuur 4.3) wanneer men controleert voor de accentverdeling in de rechtercontext. De globale CMH-statistieken laten zien dat dat effect niet behouden blijft in de globale statistieken ($\text{CMH-}\chi^2 = 0.30$, $\text{df} = 2$, $p > .05$). Na inspectie van de partiële analyses is er echter wel een interactie-effect merkbaar voor de gevallen met een intermediair interval achteraan (kort interval achteraan: $\text{MH-}\chi^2 = 0.37$, $\text{df} = 1$, $p > .05$; intermediair interval achteraan: $\text{MH-}\chi^2 = 7.14$, $\text{df} = 1$, $p < .008$; lang interval achteraan: $\text{MH-}\chi^2 = 2.33$, $\text{df} = 1$, $p > .05$). Daaruit blijkt met andere woorden dat globaal genomen het effect van de accentverdeling in de linkercontext, zoals dat in

figuur 4.3 vastgesteld kon worden, opgeslorpt wordt door het effect van de accentverdeling in de rechtercontext, maar desondanks bleek er wel een effect te zijn van de accentverdeling in de linkercontext (in de verwachte richting) indien het interval aan de rechterzijde van de eindgroep min of meer ideaal is (2-3 onbeklemtoonde lettergrepen; scenario E, F en I).

Wanneer men de rollen van verklarende en controlevariabele omdraait, dan kan men constateren dat het effect van de rechtercontext op de woordvolgorde onder controle van de linkercontext wel volledig behouden blijft: $CMH-\chi^2 = 10.99$, $df = 2$, $p < .005$. De resultaten van de partiële analyses geven echter te kennen dat het effect van accentverdeling in de rechtercontext niet voor alle partiële tabellen geldt, maar slechts enkel voor de gevallen waar er aan de linkerzijde een accentenbotsing is (kort interval vooraan: $MH-\chi^2 = 7.14$, $df = 1$, $p < .008$; intermediair interval vooraan: $MH-\chi^2 = 0.76$, $df = 1$, $p > .05$; lang interval vooraan: $MH-\chi^2 = 2.81$, $df = 1$, $p > .05$; scenario B, D en F).

Samenvattend kunnen we zeggen dat het beeld dat uit de conditionele analyses naar voren is gekomen, de effecten van de accentverdeling in de linker- en rechtercontext danig modificeert: er is wel degelijk sprake van een effect van de linkercontext op de keuze voor de woordvolgorde, maar enkel in de gevallen waar het interval met onbeklemtoonde lettergrepen aan de rechterkant ideaal is; in gevallen waar een te groot interval of een te klein interval (accentenbotsing) in de rechtercontext dreigt, heeft de accentverdeling in de linkercontext geen effect. De accentverdeling in de rechtercontext heeft slechts een effect, indien er in de linkercontext een te kort interval met onbeklemtoonde lettergrepen is; in de andere gevallen heeft het geen effect op de keuze voor de woordvolgorde.

4.6 Samenvatting

In dit hoofdstuk hebben we de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep met een deelwoord en een hulpwerkwoord vanuit een prosodisch perspectief bestudeerd. De analyses waarover in dit hoofdstuk gerapporteerd is, hebben duidelijk gemaakt dat de keuze voor de groene [PART+AUX] of rode [AUX+PART] volgorde beïnvloed wordt door de

accentdistributie voor, in en na de werkwoordelijke eindgroep. Vooraleer we voor de drie contexten de belangrijkste resultaten samenvatten, wijzen we er eerst op dat de overkoepelende hypothese (het zgn. accentueringsprincipe of de accentueringshypothese) die in dit hoofdstuk getest is, de volgende componenten bevat: in een Nederlandse zin wordt er enerzijds naar gestreefd om elk accent vooraf te laten gaan en te laten volgen door 2-3 onbeklemtoonde syllabes, en anderzijds om het platte-hoedpatroon te realiseren, i.e. het einde van een ‘normale’ zin moet gemarkeerd worden door een lage grenstoon (een onbeklemtoonde syllabe). Indien een accentenbotsing dreigt, indien een interval met onbeklemtoonde lettergrepen te groot wordt of indien de zin niet met een onbeklemtoonde syllabe eindigt, kan de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep, met het onbeklemtoonde hulpwerkwoord en het beklemtoonde deelwoord, als een relatief correctiemechanisme gebruikt worden om een accentenbotsing of een te groot interval te vermijden of een lage grenstoon te realiseren.

4.6.1 De accentdistributie in de werkwoordelijke eindgroep

Het onderzoek naar de invloed van de accentuering binnen de werkwoordelijke eindgroep hebben we uitgevoerd aan de hand van een indirecte operationalisering (scheidbaar samengesteld versus niet-scheidbaar samengesteld deelwoord) en twee directe operationalisering (de afstand van het deelwoordaccent tot het begin van de eindgroep en tot het eind van de eindgroep). De analyses hebben in de eerste plaats duidelijk gemaakt dat er een effect is van de morfologische structuur van het deelwoord op de keuze voor een van de woordvolgordevarianten, en wel in die zin dat de voorkeur voor de rode volgorde [AUX+PART] groter is als het deelwoord scheidbaar samengesteld is. Dat resultaat ligt volledig in de lijn van het accentueringsprincipe, althans wanneer men aanneemt dat er vlak voor de werkwoordelijke eindgroep een zwaar beklemtoond element staat: door voor de rode volgorde [AUX+PART] te kiezen, met het onbeklemtoonde hulpwerkwoord tussen het preverbale accent en het deelwoordaccent, kan een accentenbotsing vermeden worden. De resultaten van de specifiekere analyses bevestigden eveneens het bestaan van het accentueringsprincipe: enerzijds stijgt de keuze voor de rode volgorde [AUX+PART] naarmate er minder onbeklemtoonde syllabes tussen het woordaccent van het deelwoord en het begin van de werkwoordelijke eindgroep zitten (vermijden van

accentenbotsing door het onbeklemtoonde hulpwerkwoord tussen de twee accenten te plaatsen) en anderzijds stijgt de voorkeur voor de groene volgorde [PART+AUX] naarmate er minder onbeklemtoonde syllabes tussen het woordaccent van het deelwoord en het einde van de werkwoordelijke eindgroep zitten (ervoor zorgen dat de lage grenssteen gerealiseerd kan worden door de eindgroep af te sluiten met het onbeklemtoonde hulpwerkwoord).

Bij een vergelijking van de directe operationaliseringen met de indirecte konden we tot slot nog vaststellen dat de indirecte operationalisering grotendeels hetzelfde effect vat als de indirecte, waaruit besloten kan worden dat de operationalisering in termen van morfologische structuren voor de huidige doelstellingen accuraat is.

4.6.2 De accentdistributie in de linkercontext

Het onderzoek naar de invloed van de accentuering in de linkercontext van de werkwoordelijke eindgroep hebben we uitgevoerd aan de hand van een aantal indirecte operationaliseringen (definiete vs. indefiniete substantiefgroepen, graad van informationaliteit, aan- vs. afwezigheid van een primaire bepaling en lengte van het middenstuk) en een directe operationalisering (de afstand van het deelwoordaccent tot het laatste preverbale woordaccent). De verschillende analyses hebben in de eerste plaats duidelijk gemaakt dat er een effect is van alle directe en indirecte operationaliseringen. Zo bevordert de aanwezigheid van een indefiniete substantiefgroep, een hoog-informationeel woord en een primaire bepaling in de onmiddellijke linkercontext van de eindgroep de rode volgorde [AUX+PART]. Daarnaast kon vastgesteld worden dat de lengte van een middenstuk positief correleert met het aantal rode volgordes [AUX+PART]. Die resultaten liggen volledig in de lijn van het accentueringsprincipe, althans wanneer men aanneemt dat een indefiniete substantiefgroep, een hoog-informationeel woord, een primaire bepaling en een lang middenstuk een grotere kans hebben om geaccentueerd te worden dan een definiete substantiefgroep, een laag-informationeel woord, een niet-primaire bepaling en een kort middenstuk. Door nl. voor de rode volgorde te kiezen [AUX+PART], met het onbeklemtoonde hulpwerkwoord tussen het preverbale accent en het deelwoordaccent, kan een accentenbotsing vermeden worden. De resultaten van de specifiek-

prosodische analyse bevestigden eveneens het bestaan van het accentueringsprincipe: de keuze voor de rode volgorde [AUX+PART] stijgt naarmate er minder onbeklemtoonde syllabes tussen het woordaccent van het deelwoord en het laatste preverbale accent zitten (vermijden van accentenbotsing door het onbeklemtoonde hulpwerkwoord tussen de twee accenten te plaatsen).

Bij een vergelijking van de directe operationaliseringen met de indirecte konden we tot slot nog vaststellen dat de effecten van de indirecte operationaliseringen niet of niet geheel teruggevoerd konden worden tot het effect van de directe operationalisering, waaruit besloten kan worden dat de indirecte operationaliseringen voor de huidige doelstellingen niet geheel accuraat zijn en dat de vastgestelde effecten van die operationaliseringen op een andere manier geïnterpreteerd zullen moeten worden dan tegen de achtergrond van het accentueringsprincipe.

4.6.3 De accentdistributie in de rechtercontext

Het onderzoek naar de invloed van de accentuering in de rechtercontext van de werkwoordelijke eindgroep hebben we uitgevoerd aan de hand van een aantal indirecte operationaliseringen (aan- vs. afwezigheid van een laatste zinsplaats, grammaticale relatie tussen de laatste zinsplaats en het hoofd, en lengte van het middenstuk) en een directe operationalisering (de afstand van het deelwoordaccent tot het eerste postverbale woordaccent). De verschillende analyses hebben in de eerste plaats duidelijk gemaakt dat er een effect is van alle directe en indirecte operationaliseringen. Zo bevordert de aanwezigheid van een laatste zinsplaats en een postverbaal complement de groene volgorde [PART+AUX]. Die resultaten liggen volledig in de lijn van het accentueringsprincipe, althans wanneer men aanneemt dat een laatste zinsplaats en een postverbaal complement een grotere kans hebben om geaccentueerd te worden dan geen laatste zinsplaats en een postverbaal adjunct. Door nl. voor de groen volgorde te kiezen [PART+AUX], met het onbeklemtoonde hulpwerkwoord tussen het postverbale accent en het deelwoordaccent in, kan een accentenbotsing vermeden worden. De resultaten van de specifiek-prosodische analyse bevestigden eveneens het bestaan van het accentueringsprincipe: de keuze voor de groene volgorde [PART+AUX] stijgt naarmate er minder onbeklemtoonde syllabes tussen het woordaccent van het deelwoord en het eerste postverbale accent zitten

(vermijden van accentenbotsing door het onbeklemtoonde hulpwerkwoord tussen de twee accenten te plaatsen).

Bij een vergelijking van de directe operationaliseringen met de indirecte konden we tot slot nog vaststellen dat de effecten van de indirecte operationaliseringen niet of niet geheel teruggevoerd konden worden tot het effect van de directe operationalisering, waaruit besloten kan worden dat de indirecte operationaliseringen voor de huidige doelstellingen niet geheel accuraat zijn en dat de vastgestelde effecten van die operationaliseringen op een andere manier geïnterpreteerd zullen moeten worden dan tegen de achtergrond van het accentueringsprincipe.

4.6.4 Slotopmerkingen

Aangezien alle directe operationaliseringen van de accentdistributie een significant effect op de woordvolgorde vertoonden, kunnen we daar op een heel algemeen niveau de niet alledaagse conclusie aan verbinden dat ook in de geschreven taal een prosodisch mechanisme werkzaam kan zijn en dat, met andere woorden, de suggesties van Chafe (1988) en Bolinger (1975) in noot 5 op basis van empirisch onderzoek geverifieerd zijn.

In een tweetal slotanalyses zijn we nog nagegaan of de accentdistributie in de linker- en rechtercontext met elkaar in overeenstemming te brengen zijn. Totnogtoe hebben we nl. enkel de accentdistributie binnen de respectievelijke contexten bestudeerd, maar het spreekt voor zich dat de accentdistributie geen rekening houdt met syntactisch bepaalde grenzen. Uit die analyses bleek dat er weliswaar sprake is van een effect van de linkercontext op de keuze voor de woordvolgorde, maar enkel in die gevallen waar het interval met onbeklemtoonde lettergrepen in de rechtercontext ideaal is. In gevallen waar een te groot interval of een te klein interval (accentenbotsing) in de rechtercontext dreigt, heeft de accentverdeling in de linkercontext geen effect. De accentverdeling in de rechtercontext heeft daarentegen slechts een effect, indien er in de linkercontext een accentenbotsing plaatsvindt; in de andere gevallen heeft het geen effect op de keuze voor de woordvolgorde.

Noten:

1. Swerts (1998: 306-307) formuleert het anders: de voorkeur voor de groene volgorde [PART+AUX] blijft bestaan, ook onder druk van een onmiddellijk voorafgaand accent, als het deelwoord een primair accent draagt. Pas als het deelwoord een secundair accent draagt, wordt het naar achteren geduwd.
2. Behaghel formuleerde in 1932 al een soortgelijk accentueringsprincipe voor de variatie in de Duitse werkwoordelijke eindgroep: “[...] geht der Gruppe eine untonige Silbe oder untonige Silben voraus, so tritt die Nominalform an die erste Stelle; das untonige Verbum finitum erhält die zweite Stelle; es ergibt sich also der Wechsel von Unton, Hochton, Unton. Geht der Gruppe ein Hochton voraus, so erhält das Verbum finitum die Stelle unmittelbar nach diesem Hochton, es folgt die hochtonige Nominalform [...]” (citaat uit Ebert 1998: 6).
3. Bij de bespreking van de resultaten zal veelvuldig verwezen worden naar het accentueringsprincipe. Daarbij zal ervan afgezien worden om de verschillende componenten van het principe telkens weer te herhalen.
4. Vergelijk hiervoor bv. de voorbeeldzinnen in De Schutter (1996: 215) en Swerts (1998: 313): De Schutter geeft de voorkeur aan een accentloze, gesloten syllabe als grenssteen, terwijl Swerts ervan uitgaat dat ondanks een geaccentueerde laatste syllabe het platte-hoedpatroon toch nog behouden kan blijven.
5. Men kan zich de vraag stellen of het wel legitiem is om de invloed van de accentuering te onderzoeken in een medium dat nu net prosodische karakteristieken mist: de geschreven taal. Sommige onderzoekers beweren in ieder geval dat ondanks het ontbreken van die eigenschappen in het eindproduct, prosodie wel degelijk een rol kan spelen in het schrijfproces: “we monitor our writing sub-vocally, reading in an intonation, and the fact that the intonation is not actually shown and our reader is going to have to guess at it is as likely as not to escape our attention” (Bolinger 1975: 602) en “writers when they write, and readers when they read, experience auditory imagery of specific intonations, accents, pauses, rhythms, and voice qualities, even though the writing itself may show these features poorly if at all” (Chafe 1988: 397). We zullen er daarom voorlopig van uitgaan dat een invloed van prosodische eigenschappen in geschreven taal niet bij voorbaat uitgesloten is.
6. We zullen er in dit hoofdstuk van afzien om systematisch hypotheses te testen die betrekking hebben op de eis dat de afstand tussen twee beklemtoonde syllabes niet groter mag zijn dan 2-3 syllabes. We beperken ons voorlopig tot het opstellen van hypotheses met betrekking tot de eis dat een accentenbotsing vermeden moet worden.
7. De statistische analyses in deze sectie zijn door ons uitgevoerd op basis van de absolute cijfers die we in de geciteerde werken gevonden hebben.
8. Het is niet geheel duidelijk waarop taalbeleidsmakers zich baseren om het niet-verbale deel soms wel en soms niet aan het werkwoord te schrijven (vgl. bv. *aandoen* vs. *aan kunnen*). Het is niet ondenkbaar dat men zich daarbij laat leiden door intuïties over het al dan niet bereikte eindstadium

van een proces van grammaticalisatie of lexicalisatie en, daar nauw mee verweven, door de frequentie van samen voorkomen: hoe vaker twee woorden in elkaars buurt komen, hoe nauwer de semantische eenheid en hoe groter de neiging om ze aan elkaar te schrijven.

9. Booij & Van Santen merken wel op dat er verschillen kunnen zijn tussen scheidbaar samengestelde deelwoorden (zij zeggen liever *samenkoppelingen*) en woordgroepen als *op reis gaan*. Hun criteria zijn echter niet altijd toepasbaar en leiden niet altijd tot een eenduidig resultaat.
10. Aangezien de voorbeelden uit een corpus met enkel geschreven materiaal geëxtraheerd zijn, moet de toekenning van accenten aan bepaalde syllaben als louter indicatief opgevat worden.
11. Met uitzondering van de statistische analyses van Swerts (1998) zijn de statistische analyses in deze paragraaf door ons uitgevoerd op basis van de absolute cijfers die we in de geciteerde werken gevonden hebben.
12. De testitems komen respectievelijk uit de onderschikkingstest, de ontkenningstest, de puzzeltest en de onderschikkingstest. Zie paragraaf 3.2.1 en Haeseryn (1990: 91-133).
13. Merk daarbij op dat die definitie (gebaseerd op ANS 1997: 1227) niet gelijkloopt met die van De Schutter (1964: 70)—de afstand tussen onderwerp en dubbele werkwoordgroep—en De Schutter (1967: 55, 1976).
14. Omdat het vaak niet duidelijk is of iets adjectivisch dan wel adverbiaal gebruikt is, werden de adjectieven en adverbia in één groep ondergebracht (zie bv. ANS 1997: 394-395).
15. De groep met verba was dermate dun bevolkt, dat een incorporatie in de groep van nomina zich opdrong. Die laatste groep had de voorkeur vanwege het hoge informatiele gehalte van de verba.
16. Ook de term *collocatie* ontkomt niet aan terminologische en definitieve verwarring (zie bv. Evert & Krenn 2003). In dit hoofdstuk zal *collocatie* als technisch-statistische term verwijzen naar het significante, gefixeerde samen-voorkomen van twee woorden, onafhankelijk van de semantische kenmerken van het patroon. Daardoor zal de term als technisch equivalent van vaste verbinding gelden.
17. We selecteerden 100 observaties at random en selecteerden op basis van de verbindingenlijst van EGVD 1.4 (2004) en GWHN (1991) alle gevallen met een vaste werkwoordelijke verbinding. Daarbij bleken alle verbindingen minstens één inhoudswoord te bevatten (naast het werkwoord). In 4 van de 100 gevallen kwam dat inhoudswoord niet onmiddellijk voor de eindgroep: *beslissing nemen, vuur uitbreken, greep verliezen, aanslag plegen*. Daaruit volgt dat rekening gehouden moet worden met het feit dat op basis van de gekozen aanpak wellicht een aantal vaste verbindingen door de mazen van het net zullen glippen.
18. Merk op dat de term *bigram* hier licht afwijkend gebruikt is; normaliter slaat *bigram* op een sequentie van twee aanpalende elementen.
19. Opmerkingen: pluralia tantum en substantieven die enkel in gediminueerde vorm voorkomen, worden niet teruggevoerd op de minder gelede basisvorm; samengestelde woorden werden ontdaan van de flexievormen, maar werden wel als samenstelling behouden (*huisdeuren* wordt *huisdeur* en niet *huis* en *deur*).

20. Indien in een van de cellen de frequentie gelijk is aan 0, dan kan de LLR voor dat bigram niet berekend worden, omdat het natuurlijke logaritme van 0 gelijk is aan min oneindig en daardoor de LLR-score ook min oneindig is. Dat probleem hebben we trachten op te lossen door de nullen in de tabellen te vervangen door een erg klein getal: 1^{-24} . Op die manier zorgden we ervoor dat de LLR-berekening wel een interpreteerbare score opleverde (zonder dat de ingreep een beslissende invloed heeft op het feit of het bigram een statistisch significante collocatie is of niet).
21. We hebben daarvoor de verbindingslijst van EGVD 1.4 (2004) geraadpleegd, de idiomelijst achteraan in GWHN (1991) en de verbindingslijst die in GWHN (1991) gemarkeerd worden door een zgn. ‘Middeleeuwse vlag’.
22. We merken op dat een samenvatting in tabellarische vorm uiteraard altijd mogelijk is, bijvoorbeeld door met intervallen te werken of door de bigrammen op basis van de LLR-waarden op te splitsen in een groep met significante en een groep met niet-significante bigrammen (zie verderop in deze paragraaf). Een nadeel van die procedure is dan wel dat een deel van de informatie die in de continue variabele vervat zit, verloren gaat (het subtiele, graduele onderscheid tussen verschillende waarden verdwijnt indien men verschillende waarden in een groep samengooit).
23. De scheefheidsmaat uit R 2.0.1 is gebruikt, i.e. $N^{-1} \text{sd}(x)^{-3} \sum_i (x_i - \text{gemiddelde}(x))^3$.
24. Sommige speciale samenstellingen onttrekken zich aan die vrij algemene regel: endocentrische samenstellingen (*hoogleraar, stadhuis*), copulatieve samenstellingen, gelexicaliseerde woordgroepen, samenstellingen met een woordgroep als eerste lid en adjectivische samenstellingen in predicatief gebruik (als tussencategorie vermelden we hier ook afleidingen met een suffix dat op zichzelf een prosodisch woord kan vormen: *-achtig, -baar, -dom, -heid, -ling, -loos, -schap, -zaam*). De gevallen waarover hier sprake is, zijn over het algemeen vrij snel herkenbaar en zijn ook vaak opgenomen in EGVD 1.4 (2004). Bij twijfel werden de volgende werken geraadpleegd om tot een oplossing te komen: Smedts & Van Belle (1994); De Haas & Trommelen (1993: 360-361, 374-375, 425, 443); Booij (1995: 98, 115-119).
25. Dat resultaat loopt gelijk met wat Ebert (1998: 163) gevonden heeft in zijn Nürnbergse materiaal: “Wie bei den Individuen tritt die spätere Normalfolge IvF [de groene volgorde [PART+AUX], GDS] häufiger nach einem unbetonten vorangehenden Wort auf als nach einem betonten vorangehenden Wort”.
26. Voor de variabele *afstand tussen het deelwoordaccent en het laatste preverbaal accent* hebben we de waarden in drie categorieën samengevat: 0-1 onbeklemtoonde syllabes (accentenbotsing), 2-3 onbeklemtoonde syllabes (ideaal interval), >3 onbeklemtoonde syllabes (te groot interval).
27. Er moet op gewezen worden dat naar de positie na de tweede zinspool in de wetenschappelijke literatuur verwezen is met diverse termen, zoals extrapositie en vijfde zinsstuk. We zullen er in de rest van deze paragraaf van uitgaan dat die termen in deze context synoniem zijn.
28. Braecke houdt er rekening mee dat sommige types adjuncten, zoals een bijwoordelijke bepaling van doel, reden, beperking en richting, even gemakkelijk focaliseerbaar zijn als complementen. Omdat die types adjuncten zeer infrequent zijn in onze dataset ($n = 27$), zullen we er voorlopig van abstraheren.

29. Dergelijke zinsexterne PVCB's kunnen de vorm aannemen van een bovengeschikte of nevensgeschikte zin, een interjectie of parenthese. Er is geen statistisch significant effect van de vorm van de zinsexterne PVCB op de woordvolgorde en ook de lengte heeft geen invloed. Verder heeft een voorafgaande gepartitioneerde chi-kwadraatanalyse duidelijk gemaakt dat zinsexterne PVCB's niet statistisch significant verschillen van zinnen zonder postverbaal materiaal ($\chi^2 = 0.17$, $df = 1$, $p > .05$), zodat in het vervolg van deze paragraaf de zinnen zonder en de zinnen met zinsextern postverbaal materiaal in een groep samengenomen kunnen worden.
30. Een gepartitioneerde chi-kwadraatanalyse geeft te kennen dat er geen verschil is tussen de complexe laatste zinsplaatsen en de eenvoudige ($\chi^2 = 1.01$, $df = 1$, $p > .05$) wat de keuze voor de volgorde betreft, zodat in de rest van deze paragraaf de beide samengenomen kunnen worden.
31. In dat verband merkt ANS (1997: 1365-1366) op dat constituenten niet alleen achteropgeplaatst worden om ze in focus te plaatsen, maar ook om het middenstuk te verlichten, om er laag-informationele elementen te positioneren. In sommige gevallen is achteropplaatsing zelfs gegrammaticaliseerd. Uit die opsomming blijkt dat niet alle laatste zinsplaatsen focale, hoog-informationele elementen hoeven te bevatten.
32. In dat verband is het interessant om op te merken dat De Schutter (1996) relatief vaag blijft tijdens de bespreking van de laatste zinsplaats, in tegenstelling tot de andere factoren die aangedragen worden als argument voor het bestaan van een accentueringsprincipe: "In een grote meerderheid van de zinnen bevat het geëxtraponeerde zinsdeel de hoofdfocus van de zin [...]. Vooraanplaatsing van het voltooid deelwoord [groene volgorde [PART+AUX], GDS] zorgt er dus weer voor dat het primaire zinsaccent (op de NP van de geëxtraponeerde voorzetselconstituent of op een prominente constituent van de bijzin) en het voornaamste secundaire, dat op het hoofdwerkwoord van de (hoofd)zin ligt, *vrij ver* uit elkaar geplaatst worden" (onze nadruk).
33. Voor alternatieve termen: zie Storrer 2003: 4; voor de ontwikkeling van het onderzoek naar grammaticale relaties en voor fijnmazigere classificaties van grammaticale relaties: zie Somers 1984 en Storrer 2003.
34. Merk op dat de classificatorische moeilijkheden hiermee niet opgelost zijn. De criteria die hier voorgesteld worden, steunen grotendeels op beslissingen van de ANS (1997).
35. Waarom complementen vaker geëxtraponeerd worden, is niet zo eenvoudig uit te maken. De ANS (1997: 1365-1366) haalt in ieder geval vier mogelijke redenen voor achteropplaatsing aan: (i) de achteropplaatsing is gegrammaticaliseerd (i.e. er is geen andere mogelijkheid meer dan om de constituent te extraponeren), (ii) achteropplaatsing zorgt ervoor dat de constituent in focus staat, (iii) achteropplaatsing zorgt ervoor dat de constituent niet in focus staat en (iv) achteropplaatsing zorgt ervoor dat het middenstuk niet te zwaar wordt. Gegeven het feit dat een complement syntactisch, semantisch en pragmatisch het hoofd completeert en daardoor belangrijke informatie bevat, lijkt reden (iii) hier minder voor de hand te liggen, al is dat uiteraard zonder gericht onderzoek niet zonder meer te zeggen.

36. Dat zou ook de reden kunnen zijn waarom een minderheid van de laatste zinsplaatsen een preverbaal hoofd heeft ($n = 151$, 17.36%; het verschil in distributie met de verbale hoofden is significant: $\chi^2 = 369.53$, $df = 1$, $p < .0001$).
37. Voor de CMH-analyses is gebruik gemaakt van de operationalisering waarbij het aantal onbeklemtoonde lettergrepen tussen het woordaccent van het deelwoord en het woordaccent van het eerste postverbale inhoudswoord geteld werd. Dezelfde categorieën als in tabel 4.20 werden overgenomen: 0-1 versus 2 versus 3 versus 4 versus meer dan 4 onbeklemtoonde tussenliggende syllabe(n). In tabel 4.21 wordt daarnaar verwezen met respectievelijk A, B, C, D en E.
38. Dat hier sprake is van twee effecten (de accentverdeling in de linkercontext en rechtercontext) in plaats van drie (de accentverdeling in de eindgroep zelf), betekent niet dat we het effect van de accentverdeling in de eindgroep onbelangrijk achten of weg willen moffelen: in de analyse van de accentverdeling in de linkercontext en die in de rechtercontext is de accentverdeling in de werkwoordelijke eindgroep nl. geïncorporeerd (omdat net de afstand tussen het vorige accent en het deelwoordaccent en de afstand tussen het deelwoordaccent en het volgende accent berekend is).
39. ‘/’ hoeft aan de rechterzijde van de eindgroep niet per se te staan voor een beklemtoonde lettergreep, maar kan ook een zinsgrens representeren.

Hoofdstuk 5

De semantische dimensie

5.1 Inleiding

Bij de voorstelling van het onderzoeksobject in hoofdstuk 1 werd opgemerkt dat de interpretatie van de volgordealternatie bemoeilijkt wordt doordat een van de alternerende elementen, m.n. het hulpwerkwoord, semantisch zo goed als leeg is en een interpretatie in semantische termen daardoor op de helling komt te staan. Immers, indien twee elementen, waarvan er een inhoudsloos is, van plaats verwisselen, dan kan niet ingezien worden hoe dat een betekenisverandering teweeg zou kunnen brengen. Desalniettemin hebben tal van onderzoekers getracht de volgordevariatie in de tweeledige eindgroep in verband te brengen met de (aspectuele) betekenis van de eindgroep en de zin. Uit de bespreking van de algemene literatuur in hoofdstuk 1 bleek alvast dat bij een bepaald type tweeledige eindgroep de keuze voor de woordvolgorde op een drastische manier bepaald lijkt te worden door de status van het deelwoord: indien het deelwoord een adjectivische betekenis heeft, dan is enkel de groene volgorde [PART+AUX] mogelijk, heeft het deelwoord daarentegen een werkwoordelijke betekenis, dan kan het zowel in de groene [PART+AUX] als in de rode volgorde [AUX+PART] verschijnen. Dat de semantische status van het deelwoord zo een duidelijk en categorisch effect lijkt te hebben, heeft niet alleen de aandacht van taalwetenschappers getrokken, maar ook van taaladviseurs, waardoor een van de meest opvallende en vaakst aangehaalde regels/adviezen is ontstaan: een adjectivisch deelwoord mag nooit in de rode volgorde [AUX+PART] gebruikt worden. Taalwetenschappers hebben die regel/dat advies aan de hand van taalgebruik trachten te verifiëren, maar omdat het onderzoek ernaar gehinderd werd door

operationaliseringsproblemen—wanneer wordt een deelwoord als adjectief gebruikt en wanneer als werkwoord?—en gebrekkige hypotheses, hebben meer empirisch-georiënteerde onderzoeken de vraag naar de semantiek ook vanuit een andere hoek benaderd, met name door het verband tussen de aard van het hulpwerkwoord (*zijn* als hulpwerkwoord van tijd en passief, *zijn* als koppelwerkwoord, *hebben* en *worden* als hulpwerkwoord van tijd en passief) en de keuze voor de woordvolgorde te bestuderen.

In het huidige hoofdstuk willen we de problemen aankaarten waarmee vorig onderzoek gekampt heeft, willen we mogelijke oplossingen voor die problemen voorstellen en willen we de resultaten van dat vroegere onderzoek verifiëren. In paragraaf 5.2 zullen we daarbij in een eerste grote luik ingaan op de rol die de participiale categoriale flexibiliteit (i.e. het deelwoord als adjectief of als werkwoord?) speelt op de keuze voor de groene [PART+AUX] of de rode volgorde [AUX+PART]. In een tweede grote luik (paragraaf 5.3) wordt het effect van de aard van het hulpwerkwoord bestudeerd (dat uiteraard niet volledig losstaat van het effect van de categoriale status van het deelwoord). Paragraaf 5.4 biedt tot slot een samenvatting van de onderzoeksprestaties in dit hoofdstuk.

5.2 De impact van de categoriale flexibiliteit van het deelwoord

In deze sectie komt de rol van de categoriale status van het deelwoord aan bod. Na een eerste algemene situering van de categoriale flexibiliteit van het deelwoord (paragraaf 5.2.1), geven we een overzicht van de belangrijkste literatuur ter zake (paragraaf 5.2.2) en de hypotheses die in deze sectie getest zullen worden (paragraaf 5.2.3). In paragraaf 5.2.4 bespreken we de manier waarop we de categoriale status van het deelwoord geoperationaliseerd hebben, zodat ze in een empirische studie als deze objectief en consistent onderzocht kan worden, en in paragraaf 5.2.5 presenteren we in verschillende stappen de resultaten van de analyses. In een eerste stap (paragraaf 5.2.5.1) gaan we na wat het effect is van de categoriale status van het deelwoord, geoperationaliseerd als een categorische variabele. In een tweede stap herhalen we die analyse, waarbij de status van het deelwoord geoperationaliseerd is als continue variabele (paragraaf 5.2.5.2). Na een tussentijdse samenvatting en interpretatie (paragraaf 5.2.5.3) gaan we met behulp van een distinctieve-collexemenanalyse in een derde stap nog in op de interactie tussen de participiale en constructionele semantiek

(paragraaf 5.2.5.4). Paragraaf 5.2.6 sluit deze sectie af met een aantal slotbeschouwingen over de voorgestelde operationalisering van de status van het deelwoord.

5.2.1 *De categoriale flexibiliteit van het deelwoord: algemene situering*¹

In de traditionele grammatica gaat men ervan uit dat (sommige) deelwoorden naast verbale eigenschappen ook adjectivische eigenschappen kunnen vertonen: “Veel voltooidde deelwoorden hebben de overgang doorgemaakt naar de klasse van de adjectieven” (ANS 1997: 110; voor het Engels: zie bv. Levin & Rappaport 1986; voor het Duits: Behagel 1930 en Lenz 1993).² Die woordklasseverschuiving heeft niet alleen categoriale consequenties, maar ook semantische, doordat de oorspronkelijke werkingsbetekenis van het werkwoord naar de achtergrond verschuift en in plaats daarvan het deelwoord een bepaalde typerende eigenschap of toestand van de ermee verbonden zelfstandigheid aanduidt, soms als resultaat van de genoemde werking (verder toestandsbetekenis genoemd; ANS 1997: 387-389, Vos 1968, Versteeg 1969, Janssen 1985 & 1986).³ We illustreren met een geconstrueerd voorbeeld:

(31) Ik geloof dat de winkel al een hele tijd gesloten is.

In de voorbeeldzin is het deelwoord adjectivisch gebruikt⁴, waardoor in de zin niet zozeer een werking beschreven wordt, maar een toestand—de winkel bevindt zich in de toestand van het gesloten-zijn—en waardoor het deelwoord ook vervangen kan worden door een onverdacht adjectief, zoals *dicht*. Bij weglating van de duratieve bepaling *al een hele tijd* wordt het deelwoord, en daardoor ook de volledige zin, ambigu:

(32) Ik geloof dat de winkel gesloten is.

Indien men zoals hierboven een toestandsgerichte interpretatie aan de zin geeft, heeft dat als onmiddellijk gevolg dat het deelwoord als adjectief opgevat moet worden, indien men er een werkingsgerichte interpretatie aan geeft (waarbij *is* vervangen kan worden door het passief hulpwerkwoord *worden*), dan hebben we te maken met een verbaal deelwoord.

Verbonden met de status van het deelwoord is niet alleen de interpretatie van de volledige zin, maar ook de status van het gezegde: als het deelwoord nl. als adjectief opgevat wordt, kan het onmogelijk het hoofdwerkwoord zijn. Het hoofdwerkwoord, i.c. koppelwerkwoord, in de zin is dan de vervoegde vorm van *zijn*, met als onmiddellijk gevolg dat we met een naamwoordelijk gezegde te maken hebben. Indien het deelwoord werkwoordelijk gebruikt is, blijft het deelwoord hoofdwerkwoord van de zin en is het gezegde werkwoordelijk.⁵ In de volgende paragraaf beschrijven we de formele consequentie van de woordklasseverschuiving en de manier waarop dat geleid heeft tot een semantische verklaring van de volgordealternatie in de tweeledige werkwoordelijke eindgroep.

5.2.2 *Status quaestionis*

De implicatie van die woordklasseverschuiving voor de keuze tussen de rode [AUX+PART] en de groene volgorde [PART+AUX] is immens: traditioneel wordt aangenomen dat eindgroepen met een adjectivisch deelwoord enkel in de groene volgorde [PART+AUX] kunnen voorkomen (ANS 1997: 110, 961, 1067), net zoals ‘echte’ adjectieven enkel in de groene volgorde [PART+AUX] kunnen verschijnen (i.e. een vorm van wat we *constructionele analogie* zouden willen noemen). De consequentie van de verschuiving op formeel vlak wordt in grammatica’s meestal vergezeld van een aantal voorbeelden met een pseudodeelwoord, zoals *bereid*, dat vormelijk tot de categorie der deelwoorden gerekend kan worden, maar semantisch als adjectief fungeert (het kan niet teruggebracht worden naar een synchroon existierende infinitief):⁶

- (33) [...] dat Ciller pas bereid was tot een coalitie met Erbakan toen de beschuldigingen van corruptie opdoken. (DS)

De groene volgorde [PART+AUX] is in voorbeeld (33) voor het overgrote deel van de taaladviesliteratuur en grammatica’s de enige mogelijkheid. Dat neemt echter niet weg dat er in het taalgebruik toch voorbeelden met een pseudodeelwoord in de rode volgorde [AUX+PART] opduiken:

- (34) [...] dat de man al geruime tijd is vermist. (DS)

Dergelijke voorbeelden worden vanuit de normatieve hoek resoluut afgekeurd en als een hypercorrectie geclassificeerd (vanuit de idee dat de rode volgorde [AUX+PART] als prestigevariant gebruikt wordt in omstandigheden waar dat eigenlijk niet mogelijk is; zie daarover Sassen 1963: 18-19): “Het wil mij toeschijnen, dat mechanische toepassing van een eenmaal ‘ontdekte’ regel [...] oorzaak is van een ontsparing die zich veelvuldig voordoet [...]: men veronachtzaamt de overgang naar het adjectivische karakter, en daarmee, wat het verv. ww. [vervoegd werkwoord, GDS] betreft, de functie-overgang van hulpww. naar koppelww.” (Michels 1957: 108-109; zie verder ook nog Sassen 1963, De Rooij 1986).⁷

Moeilijker wordt het als we het terrein van de pseudodeelwoorden verlaten en kijken naar deelwoorden die zowel adjectivisch als verbaal gebruikt kunnen worden, zoals het voorbeeld met *gesloten* hierboven. In een dergelijk geval kan men zich uiteraard altijd baseren op contextueel beschikbare disambiguerende elementen (zoals de duratieve bepaling in voorbeeldzin (31)), maar die zijn lang niet altijd aanwezig. Bovendien kan men zich ook afvragen of het traditionele verband dat gezien wordt (zie bv. het citaat van Michels) tussen een adjectivisch deelwoord en een naamwoordelijk gezegde wel terecht is: ook in combinatie met *hebben*, dat traditioneel niet als koppelwerkwoord beschouwd wordt (zie echter Duinhoven 1985 & 1987), komen dergelijke adjectivische deelwoorden voor (vgl. Janssen 1985, Pardoën 1991: 1-2, voorbeeld uit Van der Horst 1993/1994: 187):

(35) Ik weet dat ze het haar opgestoken heeft.

Omgekeerd mag uit de (niet-exclusieve) band die hoe dan ook tussen een adjectivisch deelwoord en een naamwoordelijk gezegde bestaat, niet afgeleid worden dat alle naamwoordelijke gezegdes adjectivische deelwoorden bevatten, zoals de ANS (1997: 1067) lijkt te suggereren: “De plaatsing in de eindgroep van het voltooid en het passief deelwoord als zelfstandig werkwoord is in principe vrij, althans in zuiver werkwoordelijke gezegdes”. Dat er geen perfecte een-op-een-verhouding bestaat tussen adjectivische deelwoorden en naamwoordelijke gezegdes betekent dan ook dat niet alle naamwoordelijke gezegdes in de groene volgorde [PART+AUX] moeten voorkomen:

(36) [...] dat wij braaf zijn geweest. (DS)

De invloed die de status van het deelwoord op de woordvolgorde lijkt te hebben, heeft geleid tot een verklaringsmodel dat vooral door Pardoën (1991 & 1998)⁸, gebaseerd op het werk van Verhagen (1986), ontwikkeld is en door Van der Horst (1993/1994) en Duinhoven (1998) verfijnd en aangevuld. In het vervolg van deze paragraaf zullen we vooral aandacht besteden aan het basismodel van Pardoën; de verfijningen die Van der Horst en Duinhoven voorgesteld hebben, worden kort in noot 9 besproken.

Pardoën tracht de interpretatieverschillen tussen zinnen met een adjectivisch deelwoord en zinnen met een werkwoordelijk deelwoord in verband te brengen met de volgorde waarin de elementen gepresenteerd worden. Het middel dat ze daarvoor gebruikt, zijn zogenaamde interpretatiestructuren. Interpretatiestructuren geven grafisch weer hoe de woorden in een zin op elkaar betrokken moeten worden en wat de consequenties daarvan zijn op interpretatief vlak. De interpretatie van een zin, zo meent Pardoën, komt tot stand door de woorden van links naar rechts te lezen en incrementeel te interpreteren. Woorden kunnen volgens haar ofwel onmiddellijk op elkaar betrokken worden, ofwel via een tussenstapje, waarmee bedoeld wordt dat de onmiddellijke toevoeging aan het vorige even uitgesteld is. Woorden die onmiddellijk op elkaar betrokken worden, zijn niet afhankelijk van wat nog volgt in de zin, i.e. ze kunnen ten opzichte van wat nog volgt in de zin onafhankelijk geconceptualiseerd worden.

De reden waarom adjectivische deelwoorden voornamelijk in de groene volgorde [PART+AUX] lijken te voorkomen en, daaruit volgend, de reden waarom er een interpretatief verschil kan opduiken tussen de groene [PART+AUX] en de rode volgorde [AUX+PART], heeft volgens haar te maken met die onafhankelijke conceptualisering: omdat in een werkwoordelijke eindgroep in de rode volgorde [AUX+PART] het hulpwerkwoord links van het deelwoord staat en dus altijd voor het deelwoord gelezen wordt, wordt de eindgroep altijd als een eenheid geconceptualiseerd, terwijl het deelwoord in de groene volgorde [PART+AUX] daarentegen onafhankelijk van het finiete werkwoord geconceptualiseerd kan worden. Wanneer men daarbij aanneemt dat het deelwoord een a-temporele toestand aanduidt, waarbij volledig geabstraheerd wordt van het toestands- of werkingskarakter van het deelwoord, dan ligt het volgens Pardoën voor de hand dat het deelwoord in de groene

volgorde [PART+AUX], dat onafhankelijk van het finiete werkwoord geconceptualiseerd wordt, op het preverbale materiaal betrokken wordt en daardoor toestandsaanduidend werkt, omdat het niet in de tijd wordt geplaatst. De toestandsaanduiding van de groene volgorde [PART+AUX] kan echter geblokkeerd worden doordat de gedachte aan een werking de enige mogelijke is, i.e. het deelwoord kan niet zinvol op de niet-verbale elementen betrokken worden. Een deelwoord in de rode volgorde [AUX+PART] kan zoals gezegd nooit los geconceptualiseerd worden van het finiete werkwoord, waardoor het steeds in de tijd wordt geplaatst en de gedachte aan de werking allesoverheersend wordt.⁹

Hoe verhelderend de interpretatiestructuren van Pardoën ook zijn en hoe interessant de hypothese ook is dat het interpretatieverschil tussen de groene [PART+AUX] en de rode volgorde [AUX+PART] begrepen moet worden tegen de achtergrond van een andere conceptualisering, toch brengen haar post-hoc-analyses—de interpretatiestructuren ontstaan op basis van een interpretatie van de zin—weinig concreets bij aan de studie van de groene [PART+AUX] en de rode woordvolgorde [AUX+PART], precies omdat ze post-hoc zijn en in hevige mate een beroep doen op introspectie. Zo kunnen haar analyses, bijvoorbeeld, niet gebruikt worden om in een concreet geval, waar disambiguerende contextuele elementen ontbreken, uit te maken of we te maken hebben met een werkwoordelijke dan wel met een adjectivische interpretatie—ambigue gevallen hebben nl. twee interpretaties en dus ook twee interpretatiestructuren—en er wordt zelfs niet aangegeven in welke gevallen de kans op een adjectivische interpretatie bevorderd of afgeremd wordt.

Het spreekt voor zich dat we in een empirische analyse als de voorliggende principieel geïnteresseerd zijn in het verband dat Pardoën schetst tussen de volgorde en de interpretatie, maar een voorwaarde daarbij is natuurlijk wel dat geen beroep gedaan wordt op de volgorde in de eindgroep om de interpretatie te bepalen, aangezien ons dat in een cirkelredenering terecht zou brengen. Met andere woorden, om in een empirische analyse op een objectieve en consistente manier gebruik te kunnen maken van de interpretatie om de keuze voor de groene [PART+AUX] of de rode volgorde [AUX+PART] te bepalen, moet de volgorde en de interpretatie van een concrete realisatie onafhankelijk van elkaar bepaald kunnen worden. Los van het post-hoc-karakter en de introspectie doet Pardoën in haar analyses nu net wel een beroep op de

volgorde om de interpretatie te bepalen: zo zegt ze bijvoorbeeld dat bij een deelwoord in de rode volgorde [AUX+PART] altijd een dynamische interpretatie op de voorgrond treedt, terwijl het er bij de groene volgorde [PART+AUX] van afhangt of het deelwoord op een zinvolle manier op het preverbale materiaal betrokken kan worden. Een enkel voorbeeld, dat hierboven al gebruikt is en hier herhaald wordt, moge duidelijk maken dat de dynamische interpretatie van de rode volgorde [AUX+PART] op zijn minst genuanceerd dient te worden:

(37) [...] dat wij braaf zijn geweest. (DS)

Het is niet in te zien hoe een werkwoord als *zijn*, een prototypisch toestandsgericht werkwoord, op een dynamische, werkingsgerichte manier geïnterpreteerd moet worden. Dat een toestandsgericht werkwoord in de rode volgorde [AUX+PART] kan voorkomen, heeft in dezen niets te maken met hypercorrectie, daar zijn althans toch geen aanwijzingen voor te vinden in de literatuur, zodat het dynamische karakter van de rode volgorde [AUX+PART], hoe plausibel ook, aan een reïnterpretatie toe is.¹⁰

5.2.3 Hypotheses

Op basis van de net besproken literatuur kunnen we de volgende hypothesen extraheren:

- [HYP 5.1] Adjectivisch gebruikte deelwoorden komen uitsluitend in de groene volgorde [PART+AUX] voor. Werkwoordelijk gebruikte deelwoorden komen zowel in de groene [PART+AUX] als in de rode [AUX+PART] volgorde voor (paragraaf 5.2.5.1 en 5.2.5.2).
- [HYP 5.2] Wanneer een werkwoordelijke eindgroep in de rode volgorde [AUX+PART] voorkomt, is de gedachte aan een werking (een dynamische interpretatie) allesoverheersend (paragraaf 5.2.5.4).

5.2.4 Operationalisering van categoriale status van het deelwoord

5.2.4.1 *Status quaestionis*

Om te verifiëren of de woordvolgorde in de tweeledige werkwoordelijke eindgroep inderdaad afhankelijk is van de status of interpretatie van het deelwoord zal het zaak zijn om op onafhankelijke gronden uit te maken welke interpretatie aan een bepaald deelwoord in een bepaalde context gegeven moet worden. Verschillende taalwetenschappers hebben in het verleden al vanuit verschillende achtergronden en met verschillende doelen gezocht naar disambiguerende criteria, meer bepaald door ervan uit te gaan dat de distributieve verschillen tussen echte adjectieven en deelwoorden ook van toepassing zijn op de adjectivische en werkwoordelijke deelwoorden. In concreto gaat het om semantische, morfologische en syntactische criteria. We geven een kort, niet-exhaustief overzicht van de criteria in de Nederlandstalige, Duitstalige en Engelstalige literatuur.

- Semantisch criterium 1: deelwoorden met een toestandsbetekenis leunen nauwer aan bij de categorie ‘adjectief’, deelwoorden met een werkings- of handelingsbetekenis leunen nauwer aan bij de categorie ‘werkwoord’ (Michels 1957, Versteeg 1969, Duinhoven 1985, Langacker 1987, Sommerfeldt 1988, Lenz 1993, Ackerman & Goldberg 1996, ANS 1997).
- Semantisch criterium 2: deelwoorden met een overdrachtelijke betekenis, leunen nauwer aan bij de categorie ‘adjectief’ (Lenz 1993, ANS 1997, Duden 1998).
- Morfologisch criterium 1: deelwoorden die een buigings-e kunnen aanhechten, leunen nauwer aan bij de categorie ‘adjectief’ (Dittmer 1982, Sommerfeldt 1988, ANS 1997, Klooster 2001, Van Eynde 2004).
- Morfologisch criterium 2: deelwoorden die gegradeerd kunnen worden, leunen nauwer aan bij de categorie ‘adjectief’ (Sommerfeldt 1988, ANS 1997, Duden 1998, Klooster 2001).
- Morfologisch criterium 3: deelwoorden die geprefigeerd kunnen worden met *on-*, leunen nauwer aan bij de categorie ‘adjectief’ (Wasow 1977, Dittmer 1982, Levin & Rappaport 1986, Sommerfeldt 1988, Lenz 1993, Klooster 2001, Van Eynde 2004).

- Morfologisch criterium 4: deelwoorden die substantiveerbaar zijn, leunen nauwer aan bij de categorie ‘adjectief’ (Duinhoven 1985, Duden 1998).
- Morfologisch criterium 5: deelwoorden met een niet-werkwoordelijke stam, leunen nauwer aan bij de categorie ‘adjectief’ (Van Eynde 2004).
- Syntactisch criterium 1: deelwoorden die in attributieve positie kunnen staan, leunen nauwer aan bij de categorie ‘adjectief’ (Dittmer 1982, Duinhoven 1985, Levin & Rappaport 1986, Sommerfeldt 1988, ANS 1997, Duden 1998, Klooster 2001).
- Syntactisch criterium 2: deelwoorden die verbonden kunnen worden met een graadaanduidende bepaling zoals *erg* of *zeer*, leunen nauwer aan bij de categorie ‘adjectief’ (Wasow 1977, ANS 1997, Klooster 2001, Van Eynde 2004).
- Syntactisch criterium 3: een duratieve bepaling stimuleert de adjectivische interpretatie van het deelwoord, een handelend voorwerp stimuleert de werkwoordelijke interpretatie (Versteeg 1969, Lenz 1993).
- Syntactisch criterium 4: indien een deelwoord onmogelijk in de rode volgorde [AUX+PART] kan voorkomen, wijst dat op het adjectivische karakter ervan; indien een deelwoord wel in de rode volgorde [AUX+PART] kan voorkomen, wijst dat op het werkwoordelijke karakter (Michels 1957, ANS 1997, Van Eynde 2004).
- Syntactisch criterium 5: indien een deelwoord een andere valentie heeft dan het ermee corresponderende werkwoord, wijst dat op het adjectivische karakter ervan (Van Eynde 2004).

Uit het overzicht moge duidelijk geworden zijn dat de disambiguering van het deelwoord om diverse redenen vaak het onderwerp geweest is van onderzoek. Ondanks de grote hoeveelheid criteria en de plausibiliteit van die criteria, kan toch nog een en ander opgemerkt worden met betrekking tot de praktische bruikbaarheid van de criteria: (i) de semantische criteria zijn gebaseerd op erg vage semantische noties (wat is een toestand, wat is een werking?) waardoor een operationalisering niet voor de hand ligt en de intuïtie van de analist in hoge mate aangesproken wordt, zodat de objectiviteit en generaliseerbaarheid van de analyse in het gedrang komt; (ii) heel wat van de morfologische en syntactische criteria meten slechts het adjectiefpotentieel en niet de adjectiefrealiteit. Dat betekent dat men op basis van die criteria wel kan nagaan

of het mogelijk is en, zo ja, hoe waarschijnlijk het is dat een bepaald deelwoord als een adjectief gebruikt wordt, maar niet hoe een deelwoord in een specifieke context gebruikt is (tenzij natuurlijk een van de elementen die in de criteria vermeld worden in de context aanwezig zou zijn).

De criteria die wel in zekere mate bruikbaar zijn om de status van het deelwoord te bepalen, zijn het vijfde morfologische criterium (heeft het deelwoord een niet-werkwoordelijke stam?) en het derde (is er een duratieve bepaling of een handelend voorwerp in de context aanwezig?) en vijfde (heeft het deelwoord een andere valentie dan het onderliggende werkwoord?) syntactische criterium. Omdat aangenomen kan worden dat we op basis van die drie criteria slechts een minderheid van de reëel geattesteerde gegevens in onze dataset zullen kunnen disambigueren—omdat de meeste reëel geattesteerde zinnen geen duratieve bepaling bevatten en de meeste deelwoorden een werkwoordelijke stam hebben en dezelfde valentie hebben als het onderliggende werkwoord—hebben we ervoor geopteerd om een algoritme te ontwikkelen dat in staat moet zijn om de categoriale status van elk reëel geattesteerd deelwoord in een welbepaalde context te bepalen. Daarbij zullen we onder meer gebruik maken van de drie net genoemde criteria. Het criterium dat om methodologische redenen niet gebruikt kan worden in deze studie is het vierde syntactische criterium (kan het deelwoord in de rode volgorde [AUX+PART] verschijnen?), aangezien men anders in een cirkelbewijsvoering terecht zou komen—de responsvariabele wordt gebruikt als criterium om de waarde van de verklarende variabele te bepalen.

5.2.4.2 *Disambiguering van de categoriale status van het deelwoord*

In wat volgt zullen we een algoritme voorstellen dat op een fijnkorrelige en objectieve manier de reële status tracht vast te stellen van elk deelwoord dat in onze dataset voorkomt met het werkwoord *zijn*.¹¹ Op basis van het resultaat, dat onder andere gebaseerd zal zijn op een aantal van de in de wetenschappelijke literatuur vaak gebruikte disambigueringstests (zie paragraaf 5.2.4.1), moet men in staat zijn om het adjectiviteitsgehalte van een bepaald deelwoord, gegeven een bepaalde context, te kunnen vaststellen. Het adjectiviteitsgehalte van een deelwoord zal daarbij niet enkel uitgedrukt worden in discrete categorieën, zoals de traditionele binaire opdeling in

adjectivische en werkwoordelijke deelwoorden (Lenz 1993: 68), maar ook aan de hand van een continue indeling, i.e. de deelwoorden zullen worden ingeschaald op een continuüm tussen extreem werkwoordelijk en extreem adjectivisch. Door de status van het deelwoord ook als een continue variabele te operationaliseren, hoeven we geen arbitraire grenzen meer aan te nemen, komen we tegemoet aan de intuïtie van een aantal linguïsten—“maar tussen een zo duidelijk nominaal en een even duidelijk verbaal geval liggen er vele die naar behoefte de ene of de andere kleur overwegend vertonen” (Michels 1957: 110; zie ook Duinhoven 1985, Lenz 1993)—en zijn we in staat om het effect van de status van het deelwoord op een meer fijnkorrelige manier en met krachtigere statistische technieken te testen.

Vooraleer het algoritme voor te stellen, is het echter noodzakelijk eerst stil te staan bij de verschillende types deelwoorden die zich in de dataset bevinden. Het gaat daarbij nl. niet alleen om deelwoorden die in sommige omstandigheden ambigu zijn tussen een werkwoordelijke en een adjectivische interpretatie, maar ook om deelwoorden die traditioneel als onproblematisch adjectivisch beschouwd worden, zoals pseudodeelwoorden—in de ruime zin van het woord—die enkel op formele gronden als deelwoord beschouwd kunnen worden (zie hoofdstuk 2 voor de gehanteerde criteria).¹² Met dat type van heterogeniteit zal uiteraard rekening gehouden moeten worden bij de opbouw van het algoritme. De classificatie die aan het algoritme ten grondslag ligt en die precies met die heterogeniteit rekening houdt, ziet er als volgt uit:

Adjectivische deelwoorden in de vorm van een:

- *Formeel deelwoord*: het zogenaamde deelwoord behoort niet (meer) tot een werkwoordelijk paradigma. De volgende subtypes kunnen worden onderscheiden:
 - *Historisch deelwoord*: een relict van een oudere taalfase, waarbij het volledige werkwoordelijke paradigma weggevallen is met uitzondering van het deelwoord dat synchroon als adjectief wordt beschouwd. Voorbeeld: *geboren*
 - *Adjectief*: een woordvorm dat enkel formeel lijkt op een deelwoord, maar synchroon, noch historisch deel uitgemaakt heeft van een verbaal

paradigma.

Voorbeeld: *belust, gezond*.

- *Semi-formeel deelwoord*: het deelwoord maakt deel uit van een verbaal paradigma, maar het heeft een of meerdere divergerende betekenissen ontwikkeld.

Voorbeeld: *gedaan in het is gedaan met de Belgen*.

- *Adjectivisch gebruikt deelwoord*: het deelwoord maakt deel uit van een verbaal paradigma, maar kan naargelang van de context naast de typisch werkwoordelijke eigenschappen van een echt deelwoord ook adjectivische eigenschappen vertonen.

Voorbeeld: *gesloten in de winkel is al een hele tijd gesloten*.

Werkwoordelijke deelwoorden met een hulpwerkwoord van tijd (zie beschrijving in ANS 1997: 954-958).

Werkwoordelijke deelwoorden met een hulpwerkwoord van het passief: (zie beschrijving in ANS 1997: 959-960).

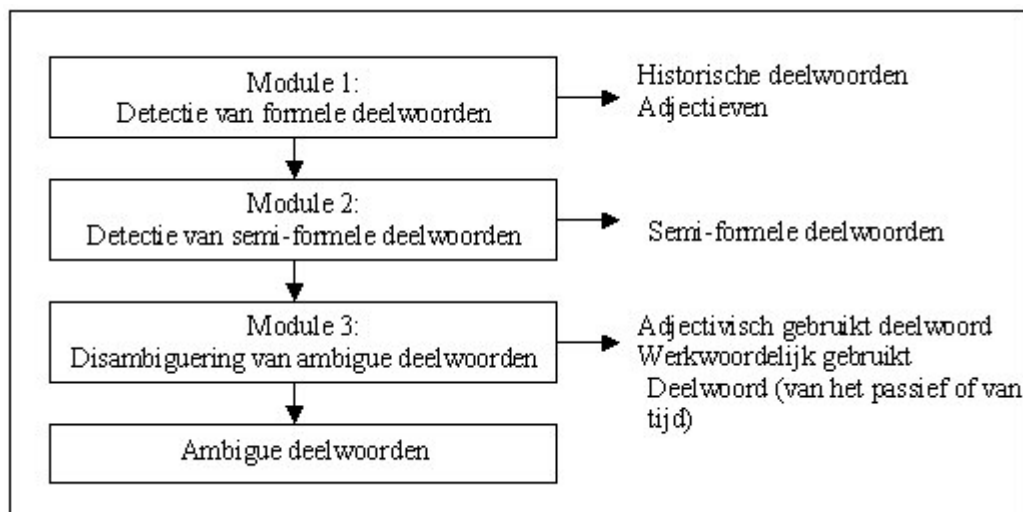
Onclassificeerbaar deelwoord: het algoritme kan geen uitsluitsel geven over de status van het deelwoord.

Architectuur van het algoritme

Het algoritme dat het deelwoord moet disambigueren bestaat uit drie modules. De eerste twee modules halen de formele en de semi-formele deelwoorden, die op basis van contextongevoelige, formele criteria gedetecteerd kunnen worden, uit de dataset. De resterende deelwoorden komen automatisch in de derde module terecht, waar zij enerzijds op basis van contextgevoelige criteria en anderzijds op basis van verschillende contextongevoelige indices gedisambiguerd zullen worden. Indien dat onmogelijk zou zijn, dan wordt het deelwoord als onclassificeerbaar beschouwd. Het algoritme is opgevat als een mechanisme dat zo snel mogelijk een deelwoord tracht te classificeren. Als het deelwoord geclassificeerd is, wordt het algoritme onmiddellijk stopgezet. Dat betekent in de praktijk dat eerst voor elk deelwoord nagegaan wordt of het een formeel deelwoord is. Indien niet, dan wordt nagegaan of het een semi-formeel

deelwoord is en zo verder. Het disambigueringsalgoritme is in die zin opgevat als een filtermechanisme waarvan de filters almaar fijnmaziger worden en waarbij de deelwoorden almaar minder duidelijk classificeerbaar zijn.

Figuur 5.1: Architectuur van het disambigueringsalgoritme.



Vorbereiding van de disambigueringsprocedure

In een eerste stap werden alle deelwoorden uit de dataset gehaald middels de filter- en exportfuncties van *Abundantia Verborum* en een aantal scripts die speciaal voor deze onderzoeksfase geschreven werden. In de dataset met 2390 observaties bleken er 417 deelwoordtypes te zijn die een tweeledige cluster vormen met een vervoegde vorm van *zijn* en derhalve in aanmerking kwamen voor disambiguering.

Module 1: detectie van de formele deelwoorden

In een eerste stap is getracht de formele deelwoorden te detecteren (historische deelwoorden en adjectieven, al dan niet in de vorm van een desubstantivische afleiding¹³) die, in lijn met de traditie, als adjectieven geclassificeerd worden (waardoor de vervoegde vorm van *zijn* als koppelwerkwoord fungeert en de cluster als werkwoordelijk deel van het naamwoordelijk gezegde dient te worden beschouwd). Van alle in onze dataset geattesteerde deelwoorden die gecombineerd zijn met een vervoegde vorm van *zijn* is in EGVD 1.4 (2004) nagegaan of er nog een link is met een synchroon existierende infinitief. Indien dat niet het geval was, werd aan de hand

van het *Woordenboek der Nederlandsche taal* (De Vries & Te Winkel 1882-1998) en De Vries et al. (1971) opgezocht of het deelwoord historisch gezien ooit tot een werkwoordelijk paradigma heeft behoord. Als er een synchrone link bestaat met een infinitief, maar de betekenis/sen van het deelwoord zo ver verwijderd is/zijn van die van de infinitief dat het onduidelijk is of beide nog op elkaar betrokken mogen worden, dan is voor dat deelwoord ook aan de hand van de net geschetste procedure nagegaan of het als formeel deelwoord beschouwd kan worden. Van de 417 deelwoordtypes bleken er volgens EGVD 1.4 (2004) 39 moeilijk of niet inpasbaar in een synchroon nog bestaand verbaal paradigma. Na raadpleging van de etymologische woordenboeken bleven er 19 formele deelwoorden over, meer bepaald 11 historische deelwoorden en 8 formeel als deelwoord uitzijnde adjectieven:¹⁴

Tabel 5.1: Overzicht van de formele deelwoordtypes in de dataset.

Formeel deelwoord: adjectief	Formeel deelwoord: Historisch deelwoord
belust < lust	bereid < bereide, rijden
bestand < bi stande	bevoegd < befugt
bevooroordeeld < vooroordeel	beschoren < bescheren
bewust < Hd. Bewusst	geboren < geberen
gedesillusioneerd < desillusie	geoorloofd < oorloven
gemeen < ghemeen	gewoon < wonen/wennen
gezond < ghesont	gezind < gezinnen
voorhanden < voor hand	herboren < herbaren
	verknocht < verknochten
	vermist < vermissen
	verschuldigd < verschuldigen

Module 2: detectie van de semi-formele deelwoorden

Om de semi-formele deelwoorden op te sporen, die gekenmerkt worden door een divergente betekenisontwikkeling, werd voor elk deelwoord in EGVD 1.4 (2004) nagegaan of ze naast de infinitief als apart lemma opgenomen zijn—vanuit de redenering dat EGVD enkel deelwoorden opneemt die betekenisdivergent zijn ten opzichte van de infinitiefvorm. Indien de betekenisomschrijving bij dat lemma overeenkwam met (een van) de betekenisomschrijving(en) van de infinitief, dan werd

het deelwoord niet als semi-formeel beschouwd. Als de betekenisomschrijving van het deelwoord echter overeenkwam met een betekenisomschrijving van de infinitief die als archaïsch of verouderd gemarkeerd staat of de betekenisomschrijving van het deelwoord kwam niet overeen met die van de infinitief, dan werd het deelwoord als semi-formeel gecategoriseerd. Het deelwoord werd eveneens als semi-formeel beschouwd als het deelwoord andere syntactische eigenschappen vertoont dan de infinitief. Een aantal voorbeelden:

- *Verkeerd*: als ‘archaïsch’ geannoteerd bij de betekenisomschrijving van de corresponderende infinitief.
- *Opgewassen*: het deelwoord heeft een specifieke betekenis die niet terug te vinden is bij de betekenisomschrijvingen van de corresponderende infinitief.
- *Gekant*: het deelwoord, in combinatie met *zijn*, heeft andere syntactische eigenschappen dan de infinitief. Het noodzakelijk aanwezige reflexief voornaamwoord ontbreekt in sommige gevallen waarin een deelwoord gecombineerd wordt met een vervoegde vorm van *zijn*.

De andere semi-formele deelwoorden zijn *aangeslagen*, *aangewezen*, *bekend*, *bezorgd*, *gedaan*, *gegrond*, *gehecht*, *geleerd*, *geschikt*, *gesteld*, *gevestigd*, *ingenomen*, *uitgeteld*, *toegedaan* en *verdeeld*.

Module 3: disambiguering van de ambigue deelwoorden

Als het deelwoord noch als formeel, noch als semi-formeel deelwoord gecategoriseerd kon worden, werd ervan uitgegaan dat het deelwoord principieel ambigu is, i.e. afhankelijk van de context als adjectivisch of als werkwoord geïnterpreteerd kan worden. Dat betekent overigens niet dat alle deelwoorden op dezelfde manier en in dezelfde mate al die eigenschappen vertonen: het is duidelijk dat een deelwoord als *gesloten* veel afhankelijker is van de context dan een deelwoord als *terechtgekomen* (dat vooral verbaal geïnterpreteerd zal worden). Het zal precies het doel zijn van deze module om daar rekening mee te houden. De module wordt zelf nog eens opgesplitst in drie consecutieve stappen (submodules) waar respectievelijk de disambiguering ondernomen zal worden op basis van contextuele criteria (adjectiefrealiteit), onafhankelijke frequentie-informatie (adjectiefpotentieel) en morfosyntactische

additietests (met als doel na te gaan of het überhaupt mogelijk is om een bepaald deelwoord als adjectivisch op te vatten; adjectieftolerantie). Voor de submodules geldt ook dat een deelwoord dat niet doorheen een filter (i.e. een van de drie submodules in het disambigueringsproces) geraakt, nooit in de volgende submodule terecht kan komen. Hieruit volgt dat de module met de contextuele criteria op een hoger niveau geplaatst wordt en in die zin belangrijker is in het beslissingsproces dan het adjectiefpotentieels- en de adjectieftolerantiemodule. Dat ligt ook voor de hand, aangezien aangenomen kan worden dat een deelwoord, los van de frequentie waarmee het in de beide interpretaties gebruikt wordt, in eerste instantie gedisambiguerd moet worden aan de hand van de context waarin het zich bevindt. Enkel en alleen wanneer die disambiguerende contextuele elementen niet aanwezig zijn, kan de frequentie van de interpretatie van belang zijn bij de disambiguering.

Vooraleer we dieper op de technische details ingaan, zal eerst een korte schets gegeven worden van de doelstellingen van de drie submodules en de manier waarop getracht zal worden de deelwoorden te disambigueren:

- Submodule A – adjectiefrealiteit: op basis van een intersubjectief gevalideerd criteriumstelsel zal een antwoord verkregen worden op de vraag welke contextuele indicatoren disambiguerend werken. Het gaat daarbij niet om de mogelijkheid om dergelijke contextuele disambiguatoren toe te voegen aan een bepaalde zin, maar over de *reële* aanwezigheid van dergelijke disambiguatoren die de *reële* adjectivische of verbale lezing van een deelwoord beïnvloeden.
- Submodule B – adjectiefpotentieel: op basis van onafhankelijk verkregen frequentie-informatie zal nagegaan worden in hoeverre een deelwoord een voorkeur heeft voor een van de beide interpretaties. Het gaat hier dus om de *mogelijkheid* om een deelwoord op een adjectivische of verbale manier te interpreteren, niet om de *reële* interpretatie. Via de frequentie-informatie trachten we enkel de waarschijnlijkheid van een bepaalde interpretatie af te wegen ten opzichte van de waarschijnlijkheid van de andere interpretatie om op die manier tot een beslissing te komen. Adjectiefpotentieel wordt op twee manieren geoperationaliseerd:
 - o Submodule B1 – indirecte-adjectiviteitsindex: op basis van frequentie-informatie in Celex (Baayen et al. 1993) en CGN (Oostdijk 2000, R5)

wordt een antwoord gezocht op de vraag of er deelwoorden zijn die proportioneel frequenter gebruikt worden dan de corresponderende werkwoordsvormen (infinitief en pv). Hoe proportioneel sterker het deelwoord staat ten opzichte van de andere werkwoordsvormen, hoe groter de kans dat het deelwoord gelexicaliseerd is als adjectief.¹⁵

- Submodule B2 – directe-adjectiviteits- en verbaliteitsindex: op basis van de frequentie-informatie in Celex (Baayen et al. 1993) en CGN (Oostdijk 2000, R5) wordt een antwoord gezocht op de vraag of en hoe vaak onafhankelijke corpusannoteerders een bepaald deelwoord als adjectivisch beschouwd hebben en hoe vaak als verbaal.
- Submodule C: op basis van informatie in EGVD 1.4 (2004) en morfosyntactische additietests zal een antwoord gezocht worden op de vraag of een bepaald deelwoord adjectiefcompatibel of adjectieftolerant is, i.e. überhaupt een adjectivische interpretatie toelaat.

Submodule A: adjectiefrealiteit

Uit het literatuuroverzicht in 5.2.4.1 is gebleken dat een aantal syntactische criteria gebruikt kan worden om duidelijkheid te verkrijgen over de status van het deelwoord.¹⁶ Zoals reeds eerder gezegd gaat het ons daarbij niet om de mogelijkheid om een bepaalde constituent aan (de context van) het te disambigueren deelwoord toe te voegen, maar wel om de reële aanwezigheid van die constituenten. In plaats van kritiekloos de bovenstaande syntactische criteria over te nemen, is eerst een kleinschalige enquête gehouden onder vijf linguïstisch getrainde predoctorale onderzoekers¹⁷ om uit te maken welke van de criteria nu precies disambiguerend werken. De enquête bestond uit 65 zinnen die een cluster met een ambigu deelwoord, gecombineerd met een vervoegde vorm van *zijn* bevatten (dus niet de formele deelwoorden). De zinnen werden at random uit de dataset getrokken¹⁸ en werden mondeling (met voldoende context) gepresenteerd in twee delen: deel 1 bestond uit 42 zinnen die een deelwoord bevatten dat geïnterpreteerd kon worden als passief deelwoord (met een hulpwerkwoord van het passief) of als adjectivisch deelwoord (met een koppelwerkwoord), deel 2 bevatte 23 zinnen met een deelwoord dat geïnterpreteerd kon worden als voltooid deelwoord (met een hulpwerkwoord van tijd) of als adjectivisch deelwoord (met een koppelwerkwoord). Sommige deelwoorden

konden zowel als voltooid deelwoord, als passief deelwoord of als adjectivisch deelwoord geïnterpreteerd worden. De volgorde van de werkwoordelijke elementen werd ongemoeid gelaten, wat resulteerde in 18 rode clusters [AUX+PART] en 24 groene clusters [PART+AUX] in het eerste deel en 18 rode clusters [AUX+PART] en 5 groene [PART+AUX] in het tweede deel. De informanten kregen de 65 zinnen allemaal in dezelfde volgorde (zie appendix D) en zonder pauze, zodat de informanten niet merkten dat er twee delen waren. De informanten werd op voorhand duidelijk gemaakt wat het onderwerp van de enquête was—is een deelwoord adjectivisch of verbaal; doordat de informanten linguïstisch geschoold waren, kon dat betrekkelijk snel gebeuren—maar er werd niets gezegd over de syntactische criteria die de status van het deelwoord kunnen disambigueren. De informanten werd gevraagd hun intuïties als taalgebruiker aan te spreken en niet op zoek te gaan naar linguïstische argumenten pro of contra een bepaalde interpretatie. De resultaten van die enquête kunnen als volgt samengevat worden:

- De graad van consensus onder de informanten bedroeg over alle zinnen heen 80.92%. Daarbij moet wel opgemerkt worden dat door het beperkte aantal informanten de minimumgrens sowieso 60% is (als je vijf informanten hebt die kunnen kiezen uit twee mogelijke interpretaties, dan betekent dat in de praktijk dat er in het slechtste geval drie informanten voor de ene interpretatie kiezen en twee voor de andere; 3 van de 5 informanten is 60%).¹⁹ De graad van consensus bedroeg in het eerste deel 78.26%, in het tweede deel 82.38%. Er tekent zich met andere woorden een licht verschil af tussen de twee delen; dat verschil kan geïnterpreteerd worden tegen de achtergrond van de unanimiteit waarmee de deelwoorden in de groene volgorde [PART+AUX] in het tweede deel beoordeeld werden (zie verderop).
- In het eerste deel werden 16 van de 42 zinnen unaniem beoordeeld, in deel 2 was dat 8 van de 22. Van de 16 consensusgevallen in deel 1 werden 12 deelwoorden verbaal geïnterpreteerd en 4 adjectivisch. In deel 2 werden alle 8 consensusgevallen verbaal geïnterpreteerd. Die resultaten lijken er dus op te wijzen dat het makkelijker is om de zuiver werkwoordelijke gevallen aan te wijzen dan de zuiver adjectivische.
- De graad van consensus bij de groene clusters [PART+AUX] in het eerste deel bedroeg 83.33%, bij de rode clusters [AUX+PART] in het eerste deel 81.11%,

bij de groene clusters [PART+AUX] in het tweede deel 88% en bij de rode clusters [AUX+PART] in het tweede deel 75.55%. In het tweede deel blijkt dus de volgorde van de werkwoorden een belangrijke rol gespeeld te hebben in de uniformiteit onder de informanten: deelwoorden in de groene volgorde [PART+AUX] werden duidelijker als verbaal geïnterpreteerd en deden blijkbaar minder twijfel rijzen. Dat is bijzonder opmerkelijk, aangezien traditioneel aangenomen wordt dat het de rode volgorde [AUX+PART] is die de verbale interpretatie van het deelwoord stimuleert, terwijl de groene volgorde [PART+AUX] de interpretatie niet onmiddellijk in een bepaalde richting dwingt. Het spreekt voor zich dat die zinnen bijzonder belangrijk zullen zijn om te ontdekken wat er nu precies voor zorgt dat in die gevallen minder twijfel heerst. We komen daar dadelijk op terug bij de individuele analyse.

- Opmerkelijk is verder nog dat eindgroepen in de rode volgorde [AUX+PART] toch vaak als adjectivisch werden beoordeeld, waaruit afgeleid kan worden dat de opvattingen in de traditionele grammatica op zijn minst genuanceerd moeten worden.²⁰

De 24 zinnen die met een volledige consensus werden beoordeeld, werden verder aan een analyse onderworpen.²¹ Er werd daarbij in het bijzonder gekeken naar de contextuele elementen die voor de consensus van de gekozen interpretatie gezorgd zouden kunnen hebben. Die analyse leverde drie onproblematische criteria op die in de submodule *adjectiefrealiteit* gebruikt zullen worden. Daarbij moet overigens opgemerkt worden dat niet alle consensusgevallen geschikt waren om er een criterium aan te onttrekken (vermoedelijk omdat de informanten gebruik gemaakt hebben van de werkwoordssemantiek en/of de woordvolgorde):

- Als de complementzin een plaatsbepaling bevat die de actie op een bepaald punt lokaliseert, waarbij het resultaat van de actie ook geldt buiten de plaats van gebeuren, dan is het deelwoord verbaal. De interpretatie van het deelwoord, en bij uitbreiding van de hele zin, wordt daarbij ingeperkt tot een punctueel gegeven en het deelwoord moet als actioneel opgevat worden, omdat de toestand na de actie of gebeurtenis niet gebonden is aan de locatie waar de actie of gebeurtenis heeft plaatsgevonden:

(38) [...] dat de gesmokkelde diamanten gestolen waren uit de diamantmijnen in Kimberley (DS).

- Als de complementzin een temporele bepaling bevat die de actie of gebeurtenis op een specifiek moment in het verleden of op een specifieke periode in het verleden (waarbij de periode duidelijk afgebakend is) situeert, dan is het deelwoord verbaal. Bepalingen die het toelaten dat de interpretatie uitgebreid wordt tot een moment of periode in het heden en/of toekomst of een periode die aanvangt in het verleden maar voortduurt tot het heden, hoeven niet per se als verbaal beschouwd te worden. Het criterium geldt enkel voor gevallen waarbij de werkwoordelijke eindgroep in de voltooid tegenwoordige tijd staat (zin(39)). Bij een voltooid verleden tijd is het namelijk altijd mogelijk om in een dergelijke context een deelwoord in adjectivische zin te interpreteren (zin(40)):

(39) [...] dat de man recent verplicht is die rekening te betalen. (DS)

(40)[...] dat hij in 1963 betrokken was bij een aanval [...]. (DS)

- Als de complementzin een handelend voorwerp bevat dat zowel oorzaak/reden als uitvoerder van de actie/handeling is, dan is het deelwoord verbaal. De agens hoeft daarbij niet per se animaat te zijn: metonymieën worden geaccepteerd.

(41) [...] dat de wapens bij hem gedeponereerd zijn door een stadsgenoot. (DS)

Submodule B: adjectiefpotentieel

Indien geen contextuele elementen aanwezig zijn die een disambiguering mogelijk maken, komen de deelwoorden in submodule B terecht waar op basis van onafhankelijke frequentie-informatie in CGN (R5) (zie hoofdstuk 2 voor een beschrijving van het corpus) en Celex (Baayen et al. 1993) nagegaan wordt hoe sterk de neiging is van een bepaald deelwoord om als adjectief of als werkwoord geïnterpreteerd te worden. Die neiging wordt uitgedrukt aan de hand van een indirecte-adjectiviteits- en een directe-adjectiviteits- en -verbaliteitsindex.

De indirecte-adjectiviteitsindex meet de graad van adjectiviteit van een deelwoord door het aantal voorkomens van het deelwoord te delen door het aantal voorkomens van het werkwoordstype (alle werkwoordsvormen van dat type):

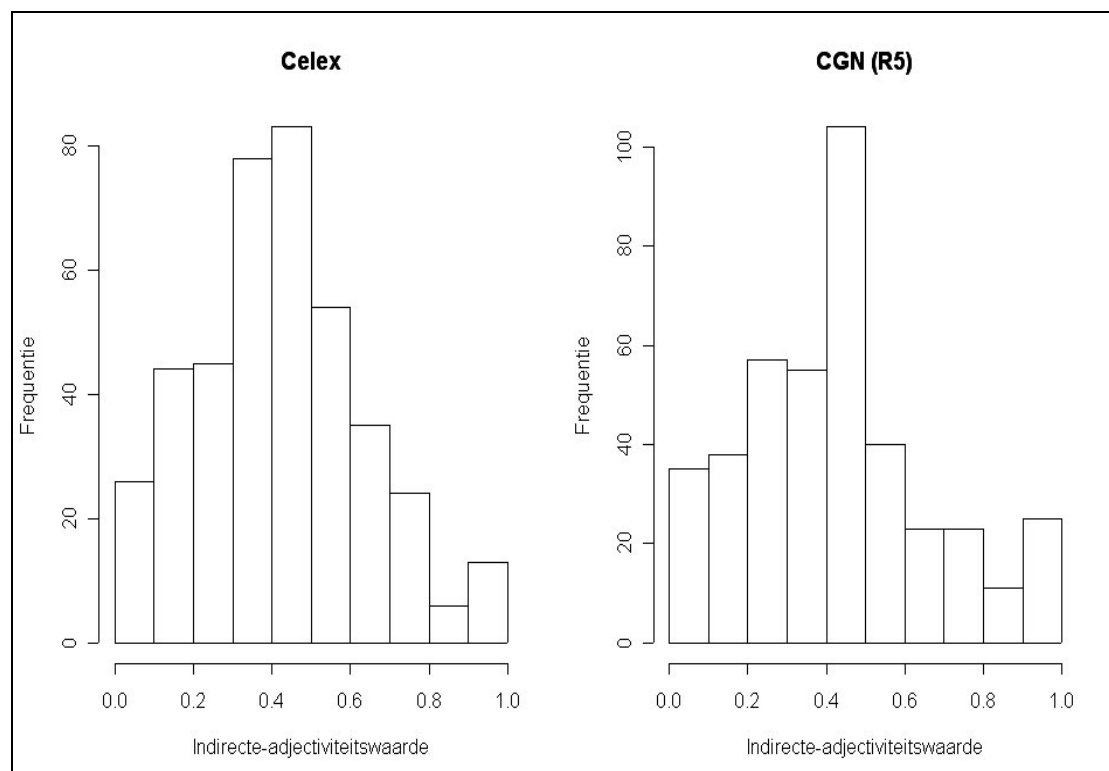
$$\text{Indirecte-adjectiviteitswaarde} = \frac{\text{Frequentie van het deelwoord } x}{\text{Frequentie van alle werkwoordsvormen van } x}$$

De uitkomst is een cijfer tussen 0 en 1: 0 betekent dat er geen participiale vormen gevonden zijn en 1 betekent dat er alleen maar participiale vormen gevonden zijn. De idee achter de indirecte-adjectiviteitsindex is dat de frequentie van het deelwoord ten opzichte van de andere werkwoordsvormen iets zegt over de mate van gespecialiseerdheid, i.e. de mate waarin naast de verbale betekenis van het deelwoord een divergente, adjectivische betekenis tot ontwikkeling gekomen is: komt het deelwoord in vergelijking met de corresponderende werkwoordsvormen vaker voor dan gemiddeld, komt het resultaat van de bovenstaande berekening met andere woorden dichterbij 1, dan kan dat een indicatie zijn dat het deelwoord een betekenisontwikkeling in de richting van de klasse der adjectieven heeft doorgemaakt, waardoor het in meer contexten gebruikt kan worden.²² In zijn meest extreme vorm is het deelwoord (als adjectief) de enige resterende vorm van het paradigma, zoals bij de formele deelwoorden, waardoor de ratio tussen het aantal voorkomens van het deelwoord en het aantal voorkomens van het volledige werkwoordstype gelijk is aan 1 (ter verduidelijking: op de formele deelwoorden passen we de maat niet toe).

Voor elk deelwoord is een indirecte-adjectiviteitswaarde berekend aan de hand van het gelemmatiseerde CGN (R5) en Celex. Het CONDIV-corpus werd om praktische redenen niet geraadpleegd, omdat het niet gelemmatiseerd is en de zoektocht naar de frequentie van de deelwoorden en de corresponderende werkwoordsvormen om die reden bijzonder tijdsintensief zou zijn. Aangezien er geen aanwijzingen zijn dat de adjectiviteitswaarde afhankelijk is van het gebruikte register en de geraadpleegde regio, is het gebruik van de frequentie-informatie uit een onafhankelijk corpus en een onafhankelijke databank geen probleem (CGN (R5) en Celex bevatten nl. meer materiaal dan enkel het formele geschreven Belgische Nederlands dat we in het onderzoek naar de taalinterne dimensies gebruiken). Met behulp van een aantal Python- en PERL-scripts werden de frequenties uit CGN (R5)

en Celex geëxtraheerd (met name de deelwoord-, de infinitief- en de pv-frequenties; de frequenties van de tegenwoordige deelwoorden werden niet in de berekening betrokken). Belangrijk om op te merken is dat bij die zoektocht om voor de hand liggende redenen geen rekening is gehouden met contextuele beperkingen (het deelwoord hoeft dus niet in een tweeledige eindgroep voor te komen) en woordsoortinformatie (het deelwoord kan met andere woorden zowel als werkwoord of als adjectief geannoteerd zijn). De lijst met werkwoordsvormen en frequenties die uit Celex geëxtraheerd werd, werd manueel gecontroleerd.²³ Vervolgens werden voor elk deelwoord twee indirecte-adjectiviteitswaarden berekend, een op basis van Celex en een op basis van CGN (R5). Om een idee te krijgen van de verdeling van de indirecte-adjectiviteitswaarden hebben we in figuur 5.2 de waarden van de indirecte-adjectiviteitsmeting (X-as) uitgezet ten opzichte van de frequentie (Y-as) waarmee ze voorkomen (we brengen in herinnering dat aan 417 deelwoordtypes een indirecte-adjectiviteitswaarde is toegekend).

Figuur 5.2: Verdeling van de indirecte-adjectiviteitswaarden.



Indien de totale frequentie van alle vormen van een werkwoord per bron kleiner is dan 5, dan werd de uitkomst genegeerd (wegens te weinig gegevens om betrouwbare uitspraken over te doen). Indien de totale frequentie hoger is dan 5, dan werden op

basis van de 2 adjectiviteitswaarden voor elk deelwoord 3 indirecte-adjectiviteitslijsten (IAL) opgesteld:

- IAL 1: bevat de deelwoorden die zowel voor CGN (R5) als voor Celex een indirecte-adjectiviteitswaarde hebben van minstens 0.66 (dat betekent dat het deelwoord minstens dubbel zo vaak voorkomt als de andere voorkomens van de werkwoordsvormen).²⁴
- IAL 2: bevat de deelwoorden die zowel voor CGN (R5) als voor Celex een indirecte-adjectiviteitswaarde hebben van minstens 0.75 (dat betekent dat het deelwoord minstens drie keer zo vaak voorkomt als de andere voorkomens van de werkwoordsvormen; zie noot 24).
- IAL 3: bevat de deelwoorden die ofwel voor CGN (R5), ofwel voor Celex een indirecte-adjectiviteitswaarde hebben van 1 (dat betekent dat enkel een deelwoord gevonden is en geen andere werkwoordsvormen).

Tabel 5.2 geeft een overzicht van de deelwoorden die aan de voorwaarden van de drie IAL's voldoen en derhalve in zekere mate beschouwd kunnen worden als vaak als adjectief gebruikte deelwoorden.

Tabel 5.2: Overzicht van de deelwoorden in de drie indirecte-adjectiviteitslijsten.

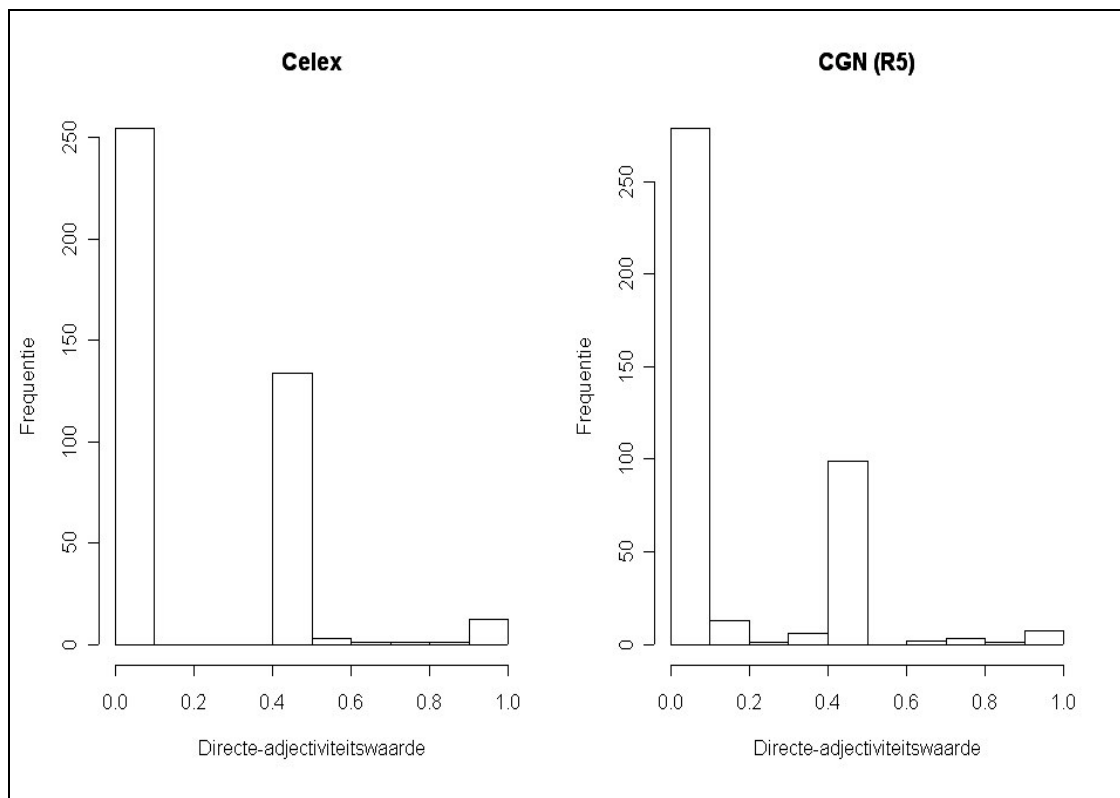
IAL 1 \geq 0.66 (n = 21)	IAL 2 \geq 0.75 (n = 8)	IAL 2 = 1 (n = 23)
afgelopen, afgesproken, afgestemd, begraven, bestemd, betrokken, bevroren, doodgeschoten, gebaseerd, gekant, genegen, geneigd, genoodzaakt, geschonden, ontslagen, opgewassen, overdreven, uitgeput, verboden, verplicht, vrijgelaten	afgelopen, begraven, bestemd, gekant, genegen, geneigd, genoodzaakt, opgewassen	aanbeland, besloten, besmet, bestreden, bewezen, gecremeerd, gekopieerd, genegen, getint, gewaarschuwd, gewezen, gezonken, gezwollen, ontdekt, ontslagen, opgewassen, uitgeroeid, verkozen, verslaafd, vertrouwd, verzadigd, voorbijgestreefd, vooringenomen

Naast de indirecte-adjectiviteitswaarde werd ook op basis van de woordsoortinformatie in CGN (R5) en Celex een directe-adjectiviteitswaarde berekend. Die waarde wordt verkregen door de totale frequentie van het als adjectief geannoteerd deelwoord te delen door de totale frequentie van het deelwoord:

$$\text{Directe-adjectiviteitswaarde} = \frac{\text{Frequentie van het als adjectief geannoteerde deelwoord } x}{\text{Frequentie van alle deelwoordsvormen van } x}$$

De uitkomst varieert van 0 tot 1: 0 betekent dat er geen adjectivische vormen gevonden zijn en 1 betekent dat er alleen maar adjectivische vormen gevonden zijn; met andere woorden, hoe hoger het cijfer, hoe dominantier de adjectivische interpretatie.

De directe adjectiviteitsmeting verschilt van de indirecte adjectiviteitsmeting doordat (i) op een directe manier gebruik gemaakt wordt van de woordsoortannotatie die door de Celex- en CGN-annotatoren (R5) aangebracht is²⁵; (ii) enkel rekening gehouden wordt met de frequentie-informatie van het deelwoord (en niet met die van de corresponderende werkwoordsvormen).²⁶ Met behulp van een aantal Python- en PERL-scripts werden de frequenties uit CGN (R5) en Celex geëxtraheerd (zonder rekening te houden met de contextuele beperkingen; zie hierboven bij de indirecte-adjectiviteitsmeting). Vervolgens werden voor elk deelwoord twee directe-adjectiviteitswaarden berekend, een op basis van Celex en een op basis van CGN (R5). Om een idee te krijgen van de verdeling van de directe-adjectiviteitswaarden hebben we in figuur 5.3 de waarden van de directe-adjectiviteitsmeting (X-as) uitgezet ten opzichte van de frequentie (Y-as) waarmee ze voorkomen (we brengen in herinnering dat aan 417 deelwoordtypes een directe-adjectiviteitswaarde is toegekend).

Figuur 5.3: Verdeling van de directe-adjectiviteitswaarden.

Indien de totale frequentie van alle vormen van een werkwoord per bron kleiner is dan 5, dan werd de uitkomst genegeerd (wegens te weinig gegevens om betrouwbare uitspraken over te doen). Indien de totale frequentie hoger is dan 5, dan werden op basis van de twee adjectiviteitswaarden voor elk deelwoord 4 directe-adjectiviteitslijsten (DAL) opgesteld:

- DAL 1: bevat de deelwoorden die zowel voor CGN (R5) als voor Celex een directe-adjectiviteitswaarde hebben van minstens 0.66 (dat betekent dat het deelwoord minstens dubbel zo vaak als adjectief geannoteerd werd dan als werkwoord; zie noot 24).
- DAL 2: bevat de deelwoorden die zowel voor CGN (R5) als voor Celex een directe-adjectiviteitswaarde hebben van minstens 0.75 (dat betekent dat het deelwoord minstens drie keer zo vaak als adjectief geannoteerd werd dan als werkwoord; zie noot 24).
- DAL 3: bevat de deelwoorden die ofwel voor CGN (R5), ofwel voor Celex een directe-adjectiviteitswaarde hebben van 1 (dat betekent dat het deelwoord enkel als adjectief geannoteerd werd).

- DAL 4: bevat de deelwoorden die zowel voor CGN (R5) als voor Celex een directe-adjectiviteitswaarde hebben die groter is dan 0 (i.e. alle deelwoorden die minstens een keer als adjectief werden geannoteerd; zie noot 24).

Tabel 5.3 geeft een overzicht van de deelwoorden die aan de voorwaarden van de vier DAL's voldoen en derhalve in zekere mate beschouwd kunnen worden als vaak als adjectief gebruikte deelwoorden.

Tabel 5.3: Overzicht van de deelwoorden in de vier directe-adjectiviteitslijsten.

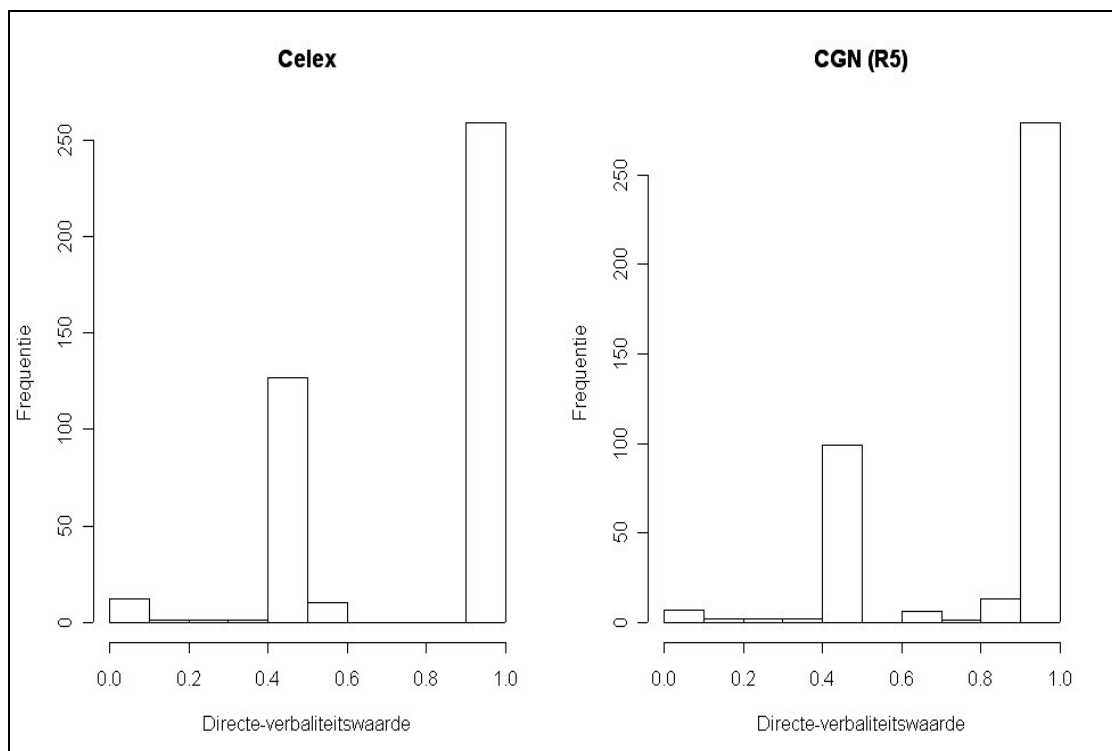
DAL 1 \geq 0.66	DAL 3 = 1 (n = 15)	DAL 4 > 0 (n = 56)
DAL 2 \geq 0.75 (n = 1)		
bekend	opgewassen, verkeerd, genegen, bekend, geoorloofd, gewezen, gewijzigd, gezegd, gezegend, gezet, gezien, gezocht, gezonken, gezwollen, vooringenomen	aangepast, afgelopen, begaan, bekend, bepaald, beperkt, beslist, besmet, betaald, betrokken, bezet, bezorgd, ervaren, gebonden, gebroken, gedaan, gedreven, gegeven, gehecht, gelaten, geleerd, genegen, geneigd, geoorloofd, gerechtvaardigd, geregeld, gericht, geschikt, geslaagd, gesloten, gesteld, getint, getrouwd, gewezen, gezien, ingenomen, ontgoocheld, opgewassen, overdreven, overtuigd, uitgeput, uitgesloten, uitgesproken, verantwoord, verboden, vergeten, verkeerd, verloren, vermoeid, verplicht, verslaafd, vertrouwd, verward, verzadigd, voldaan, voorzien

Tot slot werd ook een directe-verbaliteitswaarde berekend op basis van de woordsoortinformatie in CGN (R5) en Celex. Die waarde, die het complement is van de directe-adjectiviteitswaarde, wordt verkregen door de totale frequentie van het als werkwoord geannoteerd deelwoord te delen door de totale frequentie van het deelwoord:

$$\text{Verbaliteitswaarde} = \frac{\text{Frequentie van het als werkwoord geannoteerde deelwoord } x}{\text{Frequentie van alle deelwoordsvormen van } x}$$

Ook hier varieert de uitkomst tussen 0 tot 1: 0 betekent dat er geen verbale vormen gevonden zijn en 1 betekent dat er alleen maar verbale vormen gevonden zijn; met andere woorden, hoe hoger het cijfer, hoe dominanter de verbale interpretatie. Met behulp van een aantal Python- en PERL-scripts werden de frequenties uit CGN (R5) en Celex geëxtraheerd (zonder rekening te houden met de contextuele beperkingen; zie hierboven bij de indirecte-adjectiviteitsmeting). Vervolgens werden voor elk deelwoord twee verbaliteitswaarden berekend, een op basis van Celex en een op basis van CGN (R5). Om een idee te krijgen van de verdeling van de directe-adjectiviteitswaarden hebben we in figuur 5.4 de waarden van de directe-verbaliteitsmeting (X-as) uitgezet ten opzichte van de frequentie (Y-as) waarmee ze voorkomen (we brengen in herinnering dat aan 417 deelwoordtypes een directe-verbaliteitswaarde is toegekend).

Figuur 5.4: Verdeling van de directe-verbaliteitswaarden.



Indien de totale frequentie van alle vormen van een werkwoord per bron kleiner is dan 5, dan werd de uitkomst genegeerd (wegens te weinig gegevens om betrouwbare uitspraken over te doen). Indien de totale frequentie hoger is dan 5, dan werden op basis van de twee verbaliteitswaarden voor elk deelwoord 3 directe-verbaliteitslijsten (DVL) opgesteld:

- DVL 1: bevat de deelwoorden die zowel voor CGN (R5) als voor Celex een verbaliteitswaarde hebben van minstens 0.66 (dat betekent dat het deelwoord minstens dubbel zo vaak als werkwoord geannoteerd werd dan als adjectief; zie noot 24).
- DVL 2: bevat de deelwoorden die zowel voor CGN (R5) als voor Celex een verbaliteitswaarde hebben van minstens 0.75 (dat betekent dat het deelwoord minstens drie keer zo vaak voorkomt als werkwoord dan als adjectief; zie noot 24).
- DVL 3: bevat de deelwoorden die zowel voor CGN (R5) als voor Celex een verbaliteitswaarde hebben van 1 (dat betekent dat het deelwoord enkel als werkwoord geannoteerd is en nooit als adjectief; zie noot 24).

De lijsten met deelwoorden voor de verbaliteitsmeting worden wegens het grote aantal in appendix E opgenomen.

Submodule C: adjectieftolerantie

In de laatste submodule is tot slot nog nagegaan in hoeverre de deelwoorden in onze dataset, los van de precieze context waarin ze voorkomen, adjectiefcompatibel of – tolerant zijn, i.e. überhaupt een adjectivische interpretatie toelaten. De adjectieftolerantie hebben we in de eerste plaats getest aan de hand van een aantal bekende morfosyntactische additietests, met name prefigering door *-on*, gradering door *zeer*, *heel*, *hoogst*, *te*²⁷, prenominalisatie, comparativering (zowel synthetisch als analytisch), superlativering (zowel synthetisch als analytisch) en substantivering. Daarbij werd enkel gebruik gemaakt van CGN (R5), omdat Celex enkel informatie biedt over de woordvormen en niet over de morfosyntactische context waarin woordvormen voorkomen. Op basis van een aantal Pythonscripts werd de relevante informatie, gebruik makend van de woordsoortannotatie, uit het CGN (R5) gehaald. Daarbij zijn dialectische flexievormen buiten beschouwing gelaten. Het aantal deelwoordvormen dat positief reageerde op elk van de additietests bedraagt:

- Prefigering door *on-*: 257.
- Gradering: 58.
- Prenominalisatie: 3638.

- Comparativering of superlativering: 250.
- Substantivering: 361.

In appendix F wordt een overzicht gegeven van de deelwoordtypes die voor elk morfosyntactische test gevonden werd.

De adjectieftolerantie werd in de tweede plaats ook nog getest door na te gaan in hoeverre de betekenis waarin de deelwoorden in de dataset gebruikt zijn, overeenkomt met een van de betekenissen bij het deelwoordlemma in EGVD 1.4 (2004): zo is bijvoorbeeld voor het deelwoord *betrokken* in de dataset nagegaan of het in die vorm in EGVD 1.4 (2004) opgenomen is en zo ja, of het gebruikt is in een van de (gespecialiseerde) betekenissen die bij dat deelwoordlemma staan. Het resultaat van die zoektocht is binair: het komt overeen met een van de betekenissen in het woordenboek of er is geen deelwoordlemma / het komt niet overeen met een van de betekenissen van het deelwoordlemma.

Samenvatting van het algoritme

Tegen de achtergrond van de participiale ambiguïteit in werkwoordelijke eindgroepen met *zijn* als finiet werkwoord hebben we een algoritme opgesteld waarmee op een objectieve en eenduidige manier vastgesteld kan worden of en in welke mate een deelwoord adjectivisch of verbaal gebruikt is. Het algoritme gaat in eerste instantie op zoek naar wat we formele en semi-formele deelwoorden genoemd hebben, i.e. historische deelwoorden, adjectieven en deelwoorden met een divergente semantiek of een divergente syntactische valentie. Pas als een deelwoord niet als formeel of semi-formeel geassocieerd kan worden, gaat het op basis van een intersubjectief gevalideerd criteriumstelsel na of er zich in de context van het deelwoord een disambiguerend element bevindt (adjectiefrealiteit); indien niet, dan kan aan de hand van drie metingen, gebaseerd op de frequentie-informatie in CGN (R5) en Celex, de kans bepaald worden waarmee een bepaald deelwoord als adjectief of als werkwoord gebruikt wordt (adjectiefpotentieel). In een laatste stap wordt dan ook nog nagegaan of een deelwoord, los van de context waarin het gevonden werd, in typisch adjectivische contexten kan voorkomen (adjectieftolerantie). Om de output van het algoritme wat aanschouwelijker te maken, hebben we in tabel 5.4 voor de volgende deelwoorden in onze dataset de resultaten per module opgesomd:

- (42) [...] dat er ooit een kunstenaar met een verfkwast tegenaan was gegaan. (DS)
 (43) [...] dat de man betrokken is in een heleboel andere verdwijningen. (DS)
 (44) [...] dat zij uit een Ethiopische gevangenis waren ontsnapt [...]. (DS)

Tabel 5.4: Illustratie van de resultaten voor drie deelwoorden op basis van het disambigueringsalgoritme.

Deelwoord	Module 3A	Module 3B	Module 3C
Gegaan	-		Prenominalisatie: 0
			<i>On</i> -prefigering: 0
		IAL (Celex): 0.04	Gradering: 0
		DAL (Celex): 0	Substantivering: 0
		IAL (CGN): 0.03	Comparativering en superlativering: 0
		DAL (CGN): 0	EGVD: 0
Betrokken	-		Prenominalisatie: 0.04
			<i>On</i> -prefigering: 0
		IAL (Celex): 0.75	Gradering: 0
		DAL (Celex): 0.67	Substantivering: 0.25
		IAL (CGN): 0.70	Comparativering en superlativering: 0
		DAL (CGN): 0.15	EGVD: 1
Ontsnapt	Gedisambiguedeerd door een contextueel criterium (plaatsbepaling)		Prenominalisatie: 0
			<i>On</i> -prefigering: 0
		IAL (Celex): 0.28	Gradering: 0
		DAL (Celex): 0	Substantivering: 0
		IAL (CGN): 0.06	Comparativering en superlativering: 0
		DAL (CGN): 0.5	EGVD: 0

De manier waarop de veelheid aan resultaten per deelwoord verwerkt wordt, komt in de volgende twee secties aan bod.

Categorische classificatie van de deelwoorden

Aan het begin van paragraaf 5.2.4.2 werd aangekondigd dat de heterogene informatie die het algoritme zou leveren, gebruikt zou worden om de status van het deelwoord

zowel categorisch als continu te classificeren. De categorische classificatie vond als volgt plaats (voor de definitie van de verschillende categorieën, zie hierboven):

- Module 1: Als een deelwoord voldeed aan de voorwaarden om als formeel deelwoord beschouwd te worden, wordt het deelwoord ondergebracht in de categorie *formeel deelwoord*. Indien het niet aan de voorwaarden voldeed, dan werd getracht het deelwoord in de volgende module te disambigueren.
- Module 2: Als een deelwoord voldeed aan de voorwaarden om als semi-formeel deelwoord beschouwd te worden, wordt het deelwoord ondergebracht in de categorie *semi-formeel deelwoord*. Daarbij werd nog onderscheiden tussen de subcategorie *historisch deelwoord* en *adjectief*. Indien het niet aan de voorwaarden voldeed, dan werd getracht het deelwoord in de volgende module te disambigueren.
- Module 3 – Submodule A: Elk deelwoord dat gedisambiguerd kon worden aan de hand van de contextuele elementen werd als passief of als voltooid deelwoord gecategoriseerd.²⁸
- Module 3 – Submodule B: Voor elk deelwoord dat noch door de eerste, noch door de tweede noch door het eerste deel van de derde module gedisambiguerd kon worden, werd opgezocht in welke lijsten (DAL, IAL, DVL) het allemaal voorkomt:
 - Indien het deelwoord in minstens een van de DVL's voorkomt en niet in één van de AL's, dan werd het deelwoord als voltooid of passief deelwoord gecategoriseerd (zie noot 28).
 - Indien het deelwoord in minstens een van de IAL's en in minstens een van de DAL's voorkomt, dan werd het deelwoord als adjectivisch beschouwd.
 - Indien een deelwoord in geen van de lijsten staat, indien conflicterende informatie beschikbaar was (zowel in DVL als in AL) of indien de indexinformatie niet aan de ondergrens voldeed, dan werd het deelwoord naar submodule C gestuurd.
- Module 3 – Submodule C:
- Indien het deelwoord niet in een van de tolerantelijsten voorkomt, dan werd het als passief of voltooid deelwoord beschouwd (zie noot 28).

- Indien het deelwoord in een van de tolerantelijsten voorkomt, dan werd het deelwoord als onclassificeerbaar beschouwd, omdat we een indicatie voor adjectiviteit die enkel uit submodule C komt onvoldoende achten om een deelwoord als adjectivisch te beschouwen.
- Indien het deelwoord in een van de AL's van submodule B staat, maar niet aan de ondergrens voldeed en het deelwoord staat ook in een van de tolerantelijsten, dan werd het als adjectivisch gecategoriseerd.
- Indien het deelwoord in een van de AL's van submodule B staat, maar niet aan de ondergrens voldeed en het deelwoord staat niet in een van de tolerantelijsten, dan werd het als onclassificeerbaar beschouwd.
- Indien het deelwoord in een van de DVL's van submodule B staat, maar niet aan de ondergrens voldeed en het deelwoord staat ook in een van de tolerantelijsten, dan werd het als onclassificeerbaar beschouwd.
- Indien het deelwoord in een van de DVL's van submodule B staat, maar niet aan de ondergrens voldeed en het deelwoord staat niet in een van de tolerantelijsten, dan werd het als onclassificeerbaar beschouwd.

Continue classificatie van de deelwoorden

De operationalisering van de status van het deelwoord als een continue variabele werd op basis van de resultaten van het algoritme als volgt uitgevoerd (het algoritme werd gestopt op het ogenblik dat er een beslissing genomen was). De graad van adjectiviteit van het deelwoord wordt uitgedrukt door een cijfer tussen 0 en 1, waarbij 0 staat voor 'extreem verbaal' en 1 voor 'extreem adjectivisch':

- Module 1: Als een deelwoord voldeed aan de voorwaarden om als formeel deelwoord beschouwd te worden, werd het deelwoord de waarde '1' toegekend. Indien het niet aan de voorwaarden voldoet, dan werd getracht het deelwoord in de volgende module te disambigueren.
- Module 2: Als een deelwoord voldeed aan de voorwaarden om als semi-formeel deelwoord beschouwd te worden, werd het deelwoord de waarde '0.9' toegekend (we achten semi-formele deelwoorden iets minder duidelijk adjectivisch dan formele deelwoorden). Indien het niet aan de voorwaarden

voldoet, dan werd getracht het deelwoord in de volgende module te disambigueren.

- Module 3 – Submodule A: Elk deelwoord dat gedisambiguerd kon worden aan de hand van de contextuele elementen kreeg de waarde '0'.²⁹
- Module 3 – Submodule B en C: Elk deelwoord dat noch door de eerste, noch door de tweede noch door het eerste deel van de derde module gedisambiguerd kon worden, kreeg aan de hand van de verschillen het adjectiefpotentieels- en adjectieftolerantiemetingen een ongewogen en een gewogen gemiddelde:³⁰ het ongewogen gemiddelde wordt berekend door alle beschikbare cijfergegevens voor elk deelwoord bij elkaar op te tellen en te delen door het aantal bronnen ($n = 10$: de Celex- en CGN-waarden (R5) voor de indirecte en directe adjectiviteitsmeting en de waarden voor de vijf additietests en de EGVD-test). Het gewogen gemiddelde wordt berekend door aan het ongewogen gemiddelde een wegingsfactor toe te kennen, zodat niet alle informatie even zwaar doorweegt. Er is voor geopteerd om de waarden van het adjectiefpotentieelsmeting een wegingsfactor van 0.75 mee te geven en de waarden van de adjectieftolerantiemeting een wegingsfactor van 0.25.

5.2.5 Resultaten en discussie

5.2.5.1 *Categoriale status van het deelwoord (categorisch)*

Met het net voorgestelde algoritme zijn we nu in staat om de invloed van de status van het deelwoord, onafhankelijk en objectief vastgesteld, op de keuze voor de woordvolgorde in de tweeledige werkwoordelijke eindgroep na te gaan. Uit de status quaestionis was gebleken (i) dat traditioneel uitgegaan wordt van de onmogelijkheid om een adjectivisch deelwoord in de rode volgorde [AUX+PART] te laten voorkomen en (ii) dat er totnogtoe geen empirische analyses uitgevoerd zijn naar het effect van de adjectiviteit van het deelwoord en dat we derhalve geen echte vergelijkingsbasis hebben. Vandaar dat de hypothese die in deze sectie getest zal worden behoorlijk categorisch en in nauwelijks genuanceerde termen opgesteld is:

[HYP 5.1] Adjectivisch gebruikte deelwoorden komen uitsluitend in de groene volgorde [PART+AUX] voor. Werkwoordelijk gebruikte deelwoorden

komen zowel in de groene [PART+AUX] als in de rode [AUX+PART] volgorde voor.

In deze sectie zullen we onze analyses beperken tot de werkwoordelijke eindgroepen met een vervoegde vorm van *zijn* (n = 841; 35.19% van het totaal aantal zinnen in de dataset; zie paragraaf 4.1 voor een beschrijving van de hier gebruikte dataset), omdat daarvan voorlopig aangenomen is dat het het enige segment is binnen de dataset waarvan het deelwoord zowel een werkwoordelijke als een adjectivische interpretatie kan krijgen (zie noot 11). In paragraaf 5.3 zullen dan de eindgroepen met *hebben* en *worden* in de analyse betrokken worden.

In een eerste stap zetten we de status van het deelwoord, geoperationaliseerd als categorische variabele, uit ten opzichte van de woordvolgorde:

Tabel 5.5: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per status van het deelwoord (enkel de werkwoordelijke eindgroepen met *zijn*).

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Adjectivisch deelwoord: historisch deelwoord	94.34% (50/53)	5.66% (3/53)
Adjectivisch deelwoord: synchroon en diachroon adjectivisch	88.89% (8/9)	11.11% (1/9)
Adjectivisch deelwoord: semi-formeel deelwoord	86.11% (31/36)	13.89% (5/36)
Adjectivisch gebruikt deelwoord	84.62% (99/117)	15.38% (18/117)
Passief deelwoord	37.05% (103/278)	62.95% (175/278)
Voltooid deelwoord	23.10% (64/277)	76.90% (213/277)
Onclassificeerbaar deelwoord	56.34% (40/71)	43.66% (31/71)

Al bij een vluchtige blik op tabel 5.5 wordt duidelijk dat de adjectivische deelwoorden, conform de verwachting, een grotere voorkeur hebben voor de groene volgorde [PART+AUX] (minstens 84.62%) dan werkwoordelijke deelwoorden, die een grotere voorkeur voor de rode volgorde [AUX+PART] blijken te hebben (minstens 62.95%). De onclassificeerbare deelwoorden (i.e. deelwoorden waarvoor het algoritme niet kon uitmaken welke status het deelwoord moet krijgen) bezetten een middenpositie tussen die twee extremen. Het verschil tussen de verschillende categorieën in tabel 5.5 is statistisch significant ($\chi^2 = 171.51$, $df = 6$, $p < .0001$; $\lambda = 0.16$, $ASE = 0.02$), wat betekent dat we voor 99.99% zeker kunnen zijn dat de vastgestelde steekproefverschillen overeenkomen met reële verschillen in de volledige populatie.

Echter, in tegenstelling tot wat traditioneel aangenomen wordt, maakt tabel 5.5 eveneens duidelijk dat er op geen enkele manier sprake kan zijn van een categorisch effect van de status van het deelwoord, i.e. dat de adjectivische status van deelwoorden ervoor zou zorgen dat die deelwoorden enkel en alleen in de groene volgorde [PART+AUX] kunnen voorkomen. Uit de gegevens in onze dataset blijkt nl. dat een niet onaanzienlijke fractie adjectivische deelwoorden in de rode volgorde [AUX+PART] verschijnt (falsificatie van [HYP 5.1]) en dat, met andere woorden, zelfs voor hoogopgeleide journalisten van *De Standaard*, waarvan toch aangenomen kan worden dat ze gevoelig zijn voor en weet hebben van allerlei taalkwesties, de rode volgorde [AUX+PART], ondanks de expliciete voorschriften waarvan in hoofdstuk 1 sprake was, in dergelijke omstandigheden niet altijd een probleem vormt.³¹

Uit een specifiekere chi-kwadraattest blijkt verder nog dat de vier categorieën met adjectivische deelwoorden niet significant van elkaar verschillen ($\chi^2 = 3.22$, $df = 3$, $p > .05$). Het verschil tussen de beide types werkwoordelijke deelwoorden is daarentegen wel statistisch significant ($\chi^2 = 12.83$, $df = 1$, $p = .0003$); de o.r. van 1.96 (B.I. = 1.33-2.89) maakt daarbij ook duidelijk dat de kans op een groene volgorde [PART+AUX] versus de kans op een rode volgorde [AUX+PART] bijna twee keer zo groot is bij passieve deelwoorden dan bij voltooiden deelwoorden (op dat verschil komen we in de paragraaf 5.3 nog terug). Indien we even (om inhoudelijke redenen) zouden abstraheren van de verschillen tussen passieve en voltooiden deelwoorden en de globale categorie met adjectivische deelwoorden zouden afzetten tegen die van de

verbale deelwoorden (voltooid en passief deelwoord), dan zouden we kunnen vaststellen dat het verschil in woordvolgordevoorkeur erg significant is ($\chi^2 = 205.13$, $df = 1$, $p < .0001$) en dat de waarschijnlijkheid op een groene volgorde [PART+AUX] meer dan 16 keer groter is dan de waarschijnlijkheid op een rode volgorde [AUX+PART] als het deelwoord adjectivisch gebruikt is (o.r. = 16.11, B.I. = 10.25-26.12). Om af te sluiten hebben we ook nog een gepartitioneerde chi-kwadraatanalyse uitgevoerd, waarbij getracht wordt het aantal categorieën in tabel 5.5 te reduceren door die categorieën die niet statistisch significant van elkaar verschillen samen te nemen. De resultaten van die analyse bevestigen de eerdere resultaten—er is geen verschil tussen de types adjectivische deelwoorden, maar wel tussen de types werkwoordelijke deelwoorden; er is een verschil tussen de adjectivische en de werkwoordelijke deelwoorden—en geeft daarenboven te kennen dat ook de groep met de onclassificeerbare deelwoorden statistisch significant verschilt van de beide types werkwoordelijke deelwoorden en de groep met adjectivische deelwoorden en dat, met andere woorden de categorieën in tabel 5.5 teruggebracht kunnen worden naar vier grote categorieën, met name de adjectivische deelwoorden, de passieve deelwoorden, de voltooiden deelwoorden en de onclassificeerbare deelwoorden.

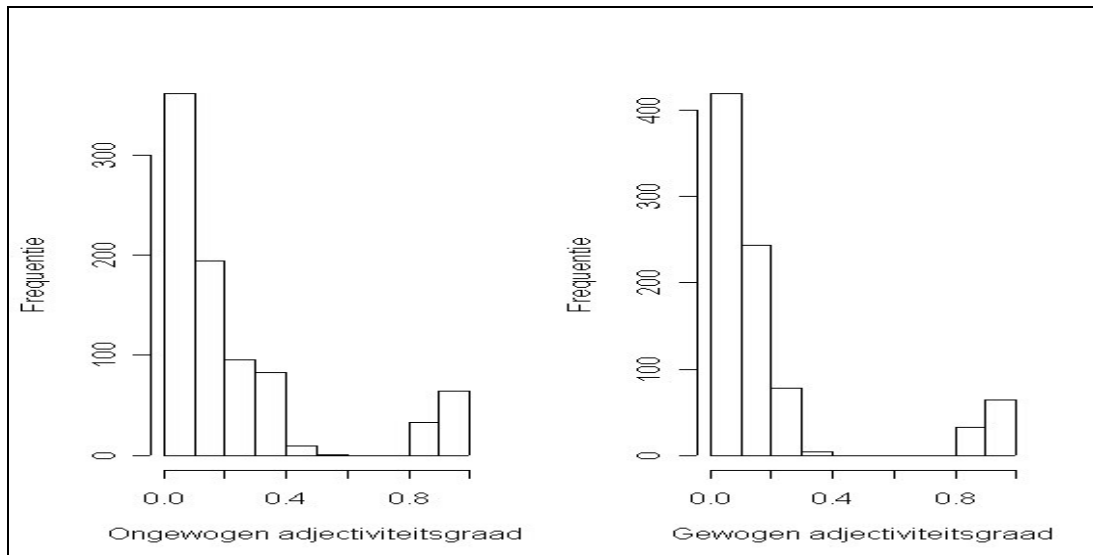
5.2.5.2 *Categoriale status van het deelwoord (continu)*

De status van het deelwoord hebben we niet alleen als een categorische variabele geoperationaliseerd, maar ook als een continue variabele, waarbij de waarden variëren tussen 0 (verbaal) en 1 (adjectivisch). Zoals hierboven al opgemerkt werd, heeft een operationalisering in de vorm van een continue verklarende variabele het voordeel dat de gegevens niet meer in discrete categorieën met arbitraire grenzen en evenwaardige leden gestopt hoeven te worden, maar dat van de subtiele gradatieverschillen in deelwoordstatus gebruik gemaakt kan worden om met krachtigere statistische technieken de invloed na te gaan op de woordvolgorde.

[HYP 5.1] Adjectivisch gebruikte deelwoorden komen uitsluitend in de groene volgorde [PART+AUX] voor. Werkwoordelijk gebruikte deelwoorden komen zowel in de groene [PART+AUX] als in de rode [AUX+PART] volgorde voor.

Gegeven [HYP 5.1] verwachten we een positieve associatie tussen de adjectiviteitswaarde en de groene volgorde [PART+AUX], i.e. hoe hoger de graad van adjectiviteit, hoe meer groene volgordes [PART+AUX]. Omdat we hier opnieuw te maken hebben met een continue variabele (vgl. de bespreking van de analyse met de LLR-waarde als verklarende variabele in hoofdstuk 4), zullen we in eerste instantie nagaan of de adjectiviteitswaarden een normale verdeling vertonen. Indien dat het geval is, kunnen we zonder meer de t-test gebruiken, anders moeten we ons, zoals in hoofdstuk 4, bedienen van de Wilcoxon-test. In figuur 5.5 worden daarom de ongewogen en de gewogen adjectiviteitsscores ten opzichte van de frequentie uitgezet.

Figuur 5.5: Verdeling van de adjectiviteitsscores.



Figuur 5.5 laat zien dat de adjectiviteitsscores niet echt normaal verdeeld lijken te zijn (42.76% van de deelwoorden hebben een adjectiviteitswaarde die zich in het interval 0-0.1 bevindt ($n = 361$), 88.36% van de waarden bevindt zich in het interval 0-0.5 ($n = 744$)). Dat de verdeling in de figuur niet-normaal is, kan aan de hand van de Shapiro-Wilktest statistisch bevestigd worden ($W = .70$, $p < .0001$: de nulhypothese voor normaliteit wordt verworpen), zodat we geen gebruik kunnen maken van de t-test en we opnieuw onze toevlucht moeten zoeken bij de Wilcoxon rank sum test om het effect te testen van de graad van adjectiviteit op de woordvolgorde. We herhalen hier kort dat, nadat de lijst met deelwoorden gerangschikt is volgens adjectiviteitswaarde en elk deelwoord een rangnummer gekregen heeft dat de positie van de adjectiviteitswaarde in de gerangschikte lijst weergeeft (de grootste adjectiviteitswaarde krijgt het rangnummer 1 mee, het tweede grootste 2, enzovoort),

de Wilcoxon-test erop gericht is na te gaan of het gemiddelde rangnummer voor de groene volgorde [PART+AUX] niet verschilt van het gemiddelde rangnummer voor de rode volgorde [AUX+PART]. Indien zou blijken dat het gemiddelde van de rangnummers voor een van de volgordevarianten significant groter of kleiner is dan de andere variant, dan kan aan de hand daarvan besloten worden dat er een verband is tussen de adjectiviteitswaarde en de keuze voor een van de volgordes. In ons geval luidt de hypothese dat voor de groep met de groene volgorde [PART+AUX] het gemiddelde rangnummer kleiner is dan voor de groep met de rode volgorde [AUX+PART] (een kleiner gemiddelde wijst erop dat de groene deelwoorden [PART+AUX] zich hogerop in de rangschikking bevinden). De resultaten van de Wilcoxon-test geven te kennen dat er inderdaad een statistisch significant effect is van de mate van adjectiviteit op woordvolgorde (ongewogen adjectiviteitsgraad: $W = 133286$, $p < .0001$; gewogen adjectiviteitsgraad: $W = 132329.5$, $p < .0001$). Om vast te stellen in welke richting dat effect dan precies geïnterpreteerd moet worden, zullen we de centrummaten in tabel 5.6 bestuderen.

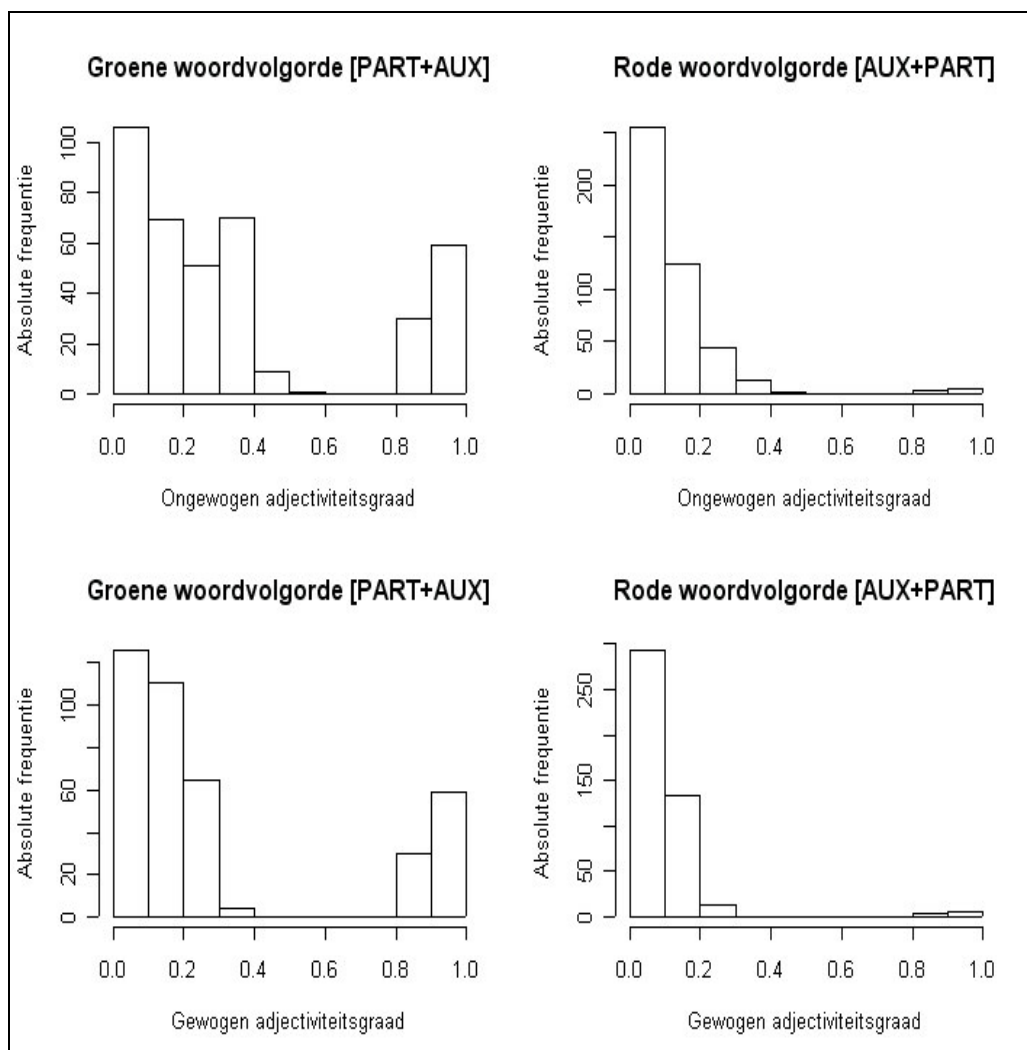
Tabel 5.6: Centrumgegevens van de gewogen en ongewogen adjectiviteitscores voor de groene [PART+AUX] en de rode [AUX+PART] woordvolgorde.

	Ongewogen adjectiviteitsgraad		Gewogen adjectiviteitsgraad	
	Groene wvo	Rode wvo	Groene wvo	Rode wvo
Gemiddelde rangnummers	133.52	64.39	119.41	59.67
Mediaan adjectiviteitsgraad	0.245	0.08	0.164	0.067
Gemiddelde adjectiviteitsgraad	0.363	0.110	0.317	0.088

In de eerste rij van tabel 5.6 worden de gemiddelde rangnummers, gesorteerd volgens adjectiviteitsgraad, voor de groene [PART+AUX] en rode [AUX+PART] woordvolgorde gegeven. Het verschil tussen de beide woordvolgordes, zowel voor de gewogen als de ongewogen adjectiviteitsmeting valt onmiddellijk op: het gemiddelde rangnummer voor de groene woordvolgorde [PART+AUX] ligt veel hoger dan het

gemiddelde rangnummer voor de rode volgorde [AUX+PART] (ongewogen adjectiviteitsgraad: 133.52 versus 64.39; gewogen adjectiviteitsgraad: 199.41 versus 59.67), hetgeen erop wijst dat de deelwoorden in de groene volgorde [PART+AUX] gemiddeld genomen een hogere adjectiviteitswaarde hebben en dus adjectivischer zijn dan de deelwoorden in de rode volgorde [AUX+PART]. De andere centrumgegevens, de gemiddelde adjectiviteitswaarde en de mediaan van de adjectiviteitswaarden, wijzen in dezelfde richting (deelwoorden in de groene volgorde [PART+AUX] hebben een hogere adjectiviteitswaarde dan deelwoorden in de rode volgorde [AUX+PART]), zodat de resultaten uit de continue analyse de verwachting bevestigen dat er een significante associatie is tussen adjectivische deelwoorden en het gebruik van de groene woordvolgorde [PART+AUX]. In figuur 5.6 wordt het verschil tussen de distributie van de adjectiviteitswaarden bij de groene volgorde [PART+AUX] (links) en bij de rode volgorde [AUX+PART] (rechts) getoond, en dit zowel voor de ongewogen meting (bovenaan) als voor de gewogen meting (onderaan). Op de Y-as van de histogrammen kan worden afgelezen hoeveel gevallen elk balkje vertegenwoordigt.

Figuur 5.6: Verdeling van de ongewogen en gewogen adjectiviteitscores per woordvolgordetype.

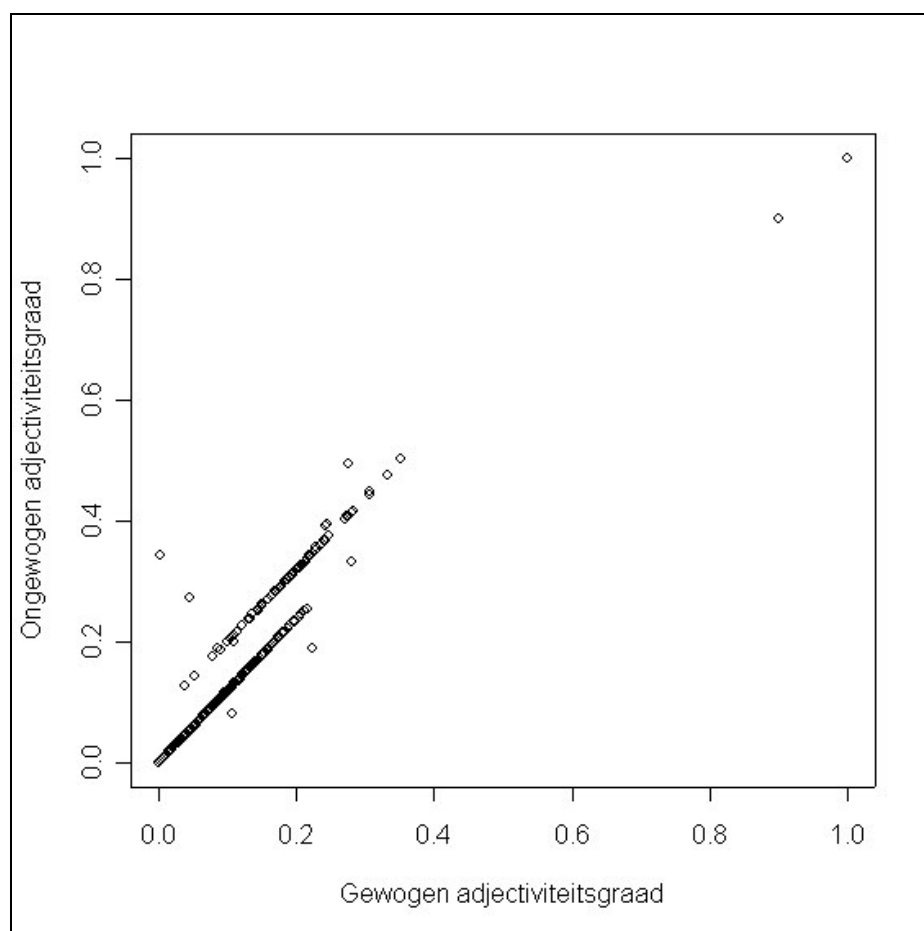


Uit de figuur komt zowel voor de ongewogen als voor de gewogen adjectiviteitsmeting *in de eerste plaats* mooi naar voren dat de deelwoorden die op basis van het disambigueringsalgoritme een hoge adjectiviteitswaarde gekregen hebben (i.e. als zeer adjectivisch beschouwd mogen worden) vooral in de groene woordvolgorde [PART+AUX] terug te vinden zijn en nauwelijks in de rode woordvolgorde [AUX+PART] (vgl. de stijging aan de rechterkant van de X-as in de grafieken met deelwoorden in de groene woordvolgorde [PART+AUX] en het ontbreken van een dergelijke stijging aan de rechterkant van de X-as in de grafieken met deelwoorden in de rode woordvolgorde [AUX+PART]). Echter, gegeven de categorische termen waarin [HYP 5.1] gesteld is, moeten we, net als bij de categorische analyse, besluiten dat er toch enkele adjectivische deelwoorden blijken te zijn die in de rode volgorde [AUX+PART] voorkomen (falsificatie van [HYP 5.1]). *In de tweede plaats* kan uit figuur 5.6 opgemaakt worden dat de deelwoorden die zich

aan de andere kant van het continuüm bevinden, met andere woorden als werkwoordelijk beschouwd worden, vaker in de rode woordvolgorde [AUX+PART] voorkomen dan in de groene woordvolgorde [PART+AUX]: de absolute frequenties aan de linkerkant van de grafiek met de deelwoorden in de groene woordvolgorde [PART+AUX] stijgen net boven de 100 uit, terwijl de absolute frequenties van de deelwoorden in de rode woordvolgorde [AUX+PART] ruim boven de 200 uitstijgen. *In de derde plaats* is de knik in de vier grafieken—vooral dan in de grafieken die de verspreiding van de deelwoorden in de groene volgorde [PART+AUX] weergeven—tussen de extreem adjectivische en de extreem werkwoordelijke gevallen (het interval 0.4/0.6 – 0.8) opvallend. Het feit dat in dat interval geen of nauwelijks deelwoorden gevonden kunnen worden, heeft alles te maken met de manier waarop in het disambigueringsalgoritme waarden aan de deelwoorden toegekend worden en zegt niets over de reële status van de deelwoorden, men kan hier met andere woorden niet uit besluiten dat deelwoorden ofwel neigen naar een extreem verbale interpretatie of een extreem adjectivische interpretatie en nauwelijks een middenpositie innemen. Als we nl. gaan kijken naar de precieze berekening van de adjectiviteitsgraad van de deelwoorden, dan kan vastgesteld worden dat de adjectiviteitswaarde van de deelwoorden die niet als formeel of semi-formeel geclassificeerd kunnen worden en niet gedisambiguerd kunnen worden door contextuele elementen, gebaseerd is op de resultaten van de adjectiviteitsmetingen én de tolerantietests, waarbij de resultaten van de zes tolerantietests, die slechts toegepast kunnen worden op een beperkt aantal deelwoorden, de gemiddelde adjectiviteitswaarde drastisch naar beneden halen (zelfs in de gewogen adjectiviteitsmeting, waar de waarden van de tolerantietests een kleinere wegingsfactor krijgen dan de adjectiviteitsmetingen). Indien we zouden abstraheren van de resultaten uit de tolerantietests en de gewogen en ongewogen berekening opnieuw zouden doen, dan zou men kunnen vaststellen dat de ongewogen adjectiviteitswaarde gemiddeld 0.14 en de gewogen adjectiviteitsmeting gemiddeld 0.18 zou stijgen (zonder de tolerantietests is de gemiddelde adjectiviteitswaarde 0.30, met de tolerantietests is de gemiddelde ongewogen waarde 0.16, de gemiddelde gewogen waarde 0.12). *In de vierde plaats* kan opgemerkt worden dat de ongewogen en gewogen adjectiviteitsmeting niet voor opvallende verschillen lijkt te hebben gezorgd, behalve dan dat op basis van de gewogen meting de breedte van het interval tussen het extreem verbale (0) en de middenpositie wat compacter is. Die constatering wordt bevestigd door figuur 5.7, waar de gewogen en ongewogen adjectiviteitsgraad

ten opzichte van elkaar uitgeplot zijn en waaruit vooral een bijna perfecte lineaire correlatie blijkt ($r = 0.986$, B.I. = 0.98-0.99). Daarnaast kan nog geconstateerd worden dat de adjectiviteitswaarden voor de verschillende deelwoorden in onze dataset zich lijken te concentreren op twee parallelle lijnen, wat erop wijst dat de deelwoorden die in de ongewogen adjectiviteitsmeting een bepaalde waarde kregen, dezelfde waarde gekregen hebben in de gewogen adjectiviteitsmeting (de onderste lijn in de grafiek doorkruist het nulpunt) of een waarde die systematisch iets hoger (in het geval van de ongewogen meting) of lager (in het geval van de gewogen meting) ligt dan de waarde in de andere meting.

Figuur 5.7: Samenhang tussen de ongewogen en gewogen adjectiviteitscores.



5.2.5.3 Tussentijdse samenvatting en interpretatie

Samenvattend kunnen we zeggen dat uit de resultaten van de categorische en continue analyses duidelijk wordt dat de keuze voor de woordvolgorde op een overweldigende manier bepaald lijkt te worden door de status van het deelwoord, waarbij een

deelwoord dat adjectivisch(er) geïnterpreteerd wordt vaker in de groene volgorde [PART+AUX] verschijnt dan een deelwoord dat werkwoordelijk(er) geïnterpreteerd wordt, al moet met klem ontkend worden dat het in het hedendaagse taalgebruik, zelfs in dat van hoogopgeleide en taalbewuste taalgebruikers, onmogelijk zou zijn om een adjectivisch deelwoord in de rode volgorde [AUX+PART] te gebruiken.

Hierboven is al opgemerkt dat de keuze voor een adjectivische dan wel werkwoordelijke interpretatie van het deelwoord rechtstreeks verband houdt met de status van het gezegde: is het deelwoord adjectivisch, dan hebben we te maken met een naamwoordelijk gezegde, is het deelwoord werkwoordelijk, dan is het gezegde werkwoordelijk. Vanuit die optiek kunnen de net gepresenteerde resultaten in verband gebracht worden met de resultaten van een van de analyses in hoofdstuk 4, met name de analyse waarin de invloed getest wordt van een naamwoordelijk deel van een naamwoordelijk gezegde (zie tabel 4.8). De resultaten van die analyse gaven, net als de resultaten van de analyses die in paragraaf 5.2.5.1 en 5.2.5.2 uitgevoerd zijn, te kennen dat de aanwezigheid van een naamwoordelijk deel (en, bij uitbreiding, de aanwezigheid van een naamwoordelijk gezegde) de keuze voor de groene volgorde [PART+AUX] significant bevordert. Uit de gelijklopendheid van de resultaten van de beide analyses mag echter in geen geval geconcludeerd worden dat de analyses volstrekt inwisselbaar zijn, aangezien ze elk op een ander type naamwoordelijk gezegde focussen: de analyse die in de vorige twee paragrafen uitgevoerd is, focust op naamwoordelijke gezegdes met een eenledig werkwoordelijk deel (de vervoegde vorm van *zijn*) en een naamwoordelijk deel in de vorm van een deelwoord, terwijl de analyse in hoofdstuk 4 focust op naamwoordelijke gezegdes met een tweeledig werkwoordelijk deel (het werkwoordelijke deelwoord *gebleken, geschenen, geleken, gedocht, gedunkt, geheten, voorgekomen, gegaan, gekomen, gelopen, geraakt, gestaan, gevallen, gezeten* en een vervoegde vorm van *zijn*) en een naamwoordelijk deel in de vorm van eender welke constituent. Bovendien is er ook een duidelijk voorkeursverschil tussen de beide types naamwoordelijke gezegdes: het type dat in deze paragraaf bestudeerd is, komt minstens in 84.62% van de gevallen in de groene volgorde [PART+AUX] voor, terwijl het type dat in hoofdstuk 4 geanalyseerd werd in slechts 69.42% van de gevallen in de groene volgorde [PART+AUX] verscheen, zodat het methodologisch en analytisch onverantwoord zou zijn om de beide types op een hoop te gooien.

De vraag die ons dan nog rest is: waarom verkiezen taalgebruikers de groene volgorde [PART+AUX] in zinnen met een naamwoordelijk gezegde en waarom gedragen niet alle naamwoordelijke gezegdes zich op een homogene manier? Ons inziens zijn er twee antwoorden mogelijk:³² (i) indien het deelwoord als adjectief gebruikt wordt, neemt het ook de syntactische valentie over van de adjectieven, wat betekent dat enkel de groene volgorde [PART+AUX] mogelijk is (een vorm van constructionele analogie). Dat hier en daar adjectivische deelwoorden toch in de rode volgorde [AUX+PART] voorkomen, heeft dan, zoals vroeger al gesuggereerd werd, te maken met een hypercorrect gebruik van de volgordevariatie: de rode volgorde [AUX+PART] is de prestigieuze variant (vgl. de resultaten in hoofdstuk 3) en komt daarom ook voor in omstandigheden waar het eigenlijk niet mogelijk is. Een dergelijke syntactische benadering verklaart meteen ook waarom het type naamwoordelijk gezegde dat in hoofdstuk 4 bestudeerd is in een veel mindere mate de groene volgorde [PART+AUX] stimuleert, m.n. omdat het deelwoord er niet adjectivisch gebruikt is, maar volledig beschikt over zijn werkwoordelijke eigenschappen. Moeilijker te verklaren is dan wel waarom dat type naamwoordelijk gezegde toch nog significant vaker in de groene volgorde [PART+AUX] voorkomt dan werkwoordelijke gezegdes. (ii) Indien het deelwoord als adjectief gebruikt wordt, neemt het de semantische eigenschappen over van de adjectieven (toestand), waardoor het volgens het model-Pardoen (zie paragraaf 5.2.2) niet in de rode volgorde [AUX+PART] kan voorkomen ([HYP 5.2]), omdat op die manier de toestandssemantiek van het deelwoord zou conflicteren met de werkingsgerichte aspectuele betekenis van de rode constructie [AUX+PART] (een deelwoord in de rode volgorde [AUX+PART] kan nooit los geconceptualiseerd worden van het finiete werkwoord, waardoor het steeds in de tijd wordt geplaatst en de gedachte aan de werking allesoverheersend wordt). Een dergelijke semantische benadering verklaart ook waarom naamwoordelijke gezegdes zonder adjectivisch deelwoord de groene volgorde [PART+AUX] significant bevorderen, met name omdat ze ook een toestandssemantiek hebben (zie ANS 1997: 1121-1122). Wat echter niet te begrijpen valt tegen de achtergrond van de semantische verklaring is waarom er een zo duidelijk voorkeursverschil is tussen de twee types naamwoordelijk gezegde: indien een naamwoordelijk gezegde als globale categorie een toestandsbetekenis heeft, dan verwacht men geen verschillen in volgordevoorkeur. Bovendien is het maar al te zeer de vraag in hoeverre het volgehouden kan worden, zoals in paragraaf 5.2.2 (zie

voorbeeldzin (37)) al aangekaart werd, dat de rode volgorde [AUX+PART] correleert met een dynamische betekenis³³, en zelfs al zou dat nog aangetoond kunnen worden, dan nog blijkt het semantische conflict tussen de constructionele en participiale semantiek op zijn minst in sommige omstandigheden al bij al wel mee te vallen, aangezien adjectivische deelwoorden sporadisch ook wel in de rode volgorde [AUX+PART] voorkomen. Toekomstig onderzoek zal zich zeker over de verhouding tussen de participiale semantiek en de constructionele semantiek moeten buigen, al zal daarvoor eerst nagedacht moeten worden over een adequate operationalisering van de werkwoordssemantiek. Precies om die reden zien we ervan af om de semantische piste hier systematisch verder uit te spitten. In de volgende paragraaf zal wel getracht worden om een eerste, voorlopige inblik te verkrijgen in de samenhang tussen de participiale en de constructionele semantiek.

5.2.5.4 *De interactie tussen de participiale en constructionele semantiek*

In deze paragraaf zal nagegaan worden in hoeverre de bevindingen van Pardoën (1991 & 1998; zie paragraaf 5.2.2) bevestigd kunnen worden. De hypothese, gebaseerd op de analyses van Pardoën, is de volgende:

[HYP 5.2] Wanneer een werkwoordelijke eindgroep in de rode volgorde [AUX+PART] voorkomt, is de gedachte aan een werking (een dynamische interpretatie) allesoverheersend.

Indien [HYP 5.2] klopt, dan verwachten we dat een deelwoord met een toestandsgerichte betekenis nooit deel kan uitmaken van een rode cluster [AUX+PART]. Anders is er sprake van een semantisch conflict tussen de deelwoordsemantiek en de constructionele semantiek.

Zoals gezegd, zullen we in deze paragraaf geen systematische semantische analyse uitvoeren, maar willen we een eerste, voorlopige idee krijgen van de interactie tussen de constructionele en de participiale semantiek. Dat kan op een betrekkelijk eenvoudige manier, meer bepaald door gebruik te maken van de collocatie-analyse die in hoofdstuk 4 geïntroduceerd werd om vaste verbindingen te detecteren. In plaats van de lexicale attractie tussen het deelwoord en een preverbaal woord te berekenen (de syntagmatische benadering), kunnen we de collocatie-analyse nl. ook gebruiken om de

lexicale attractie tussen een deelwoord en een volgordevariant te berekenen (de paradigmatische benadering, die door Gries & Stefanowitsch (2004) *distinctieve-collexemenanalyse* genoemd wordt). Op die manier krijgen we voor de beide volgordevarianten een lijstje met deelwoorden die statistisch significant vaker in die volgorde voorkomen dan in de andere volgorde. Een analyse van de werkwoorden of collexemen die distinctief zijn voor de ene of de andere constructie “may help us elucidate the existence and degree of fine semantic differences between the two [...]” (Gries & Stefanowitsch 2004: 102). De verwachting op basis van het model-Pardoen is dan dat in het lijstje met deelwoorden die significant vaker voorkomen in de rode volgorde [AUX+PART] vooral of uitsluitend werkingsgerichte deelwoorden te vinden zullen zijn, terwijl het lijstje met deelwoorden in de groene volgorde [PART+AUX] zowel werkings- als toestandsgerichte deelwoorden zal bevatten.

De berekening van de lexicale aantrekking tot een syntactische constructie zal opnieuw gebeuren met behulp van de *log-likelihood ratio test* (LLR; voor de technische details, zie hoofdstuk 4). Om de berekening succesvol te kunnen uitvoeren, hebben we de frequentiegegevens nodig van elk deelwoord in de groene volgorde [PART+AUX], elk deelwoord in de rode volgorde [AUX+PART] en de totale frequentie van de groene volgorde [PART+AUX] en de rode volgorde [AUX+PART]. In tabel 5.7 wordt in tabellarische vorm een overzicht gegeven van alle benodigde gegevens, zowel van de basisgegevens als van de afgeleide gegevens; X staat daarbij voor groene volgorde [PART+AUX], Y voor een bepaald deelwoord; \hat{X} en \hat{Y} staan voor de negatie van X en Y:

Tabel 5.7: Basisgegevens voor de berekening van de lexicale attractie met behulp van LLR

	X	\hat{X}	Totaal
Y	a	b	a+b
\hat{Y}	A-a	B-b	(A-a)+(B-b)
Totaal	A	B	A+B

De letters in de tabel staan voor de frequenties van een bepaald deelwoord in de groene volgorde [PART+AUX] (a), de frequentie van alle groene volgordes [PART+AUX] in de dataset (A), de frequentie van een bepaald deelwoord in de rode volgorde [AUX+PART] (b) en de frequentie van alle rode woordvolgordes [AUX+PART] in de dataset (B). Aan de hand van die basisgegevens kan de rest van de tabel ingevuld worden:

- (A-a) staat voor ‘de frequentie van alle deelwoorden in de groene volgorde [PART+AUX] (A), behalve die van het te onderzoeken deelwoord (a)’.
- (B-b) staat voor ‘de frequentie van alle deelwoorden in de rode volgorde [AUX+PART] (B), behalve die van het te onderzoeken deelwoord (b)’.
- (a+b) staat voor ‘de frequentie van het te onderzoeken deelwoord in de groene [PART+AUX] (a) en rode [AUX+PART] (b) woordvolgorde’.
- ((A-a)+(B-b)) staat voor ‘de totale frequentie van alle deelwoorden in de groene [PART+AUX] (A) en rode [AUX+PART] (B) woordvolgorde, behalve die van het te onderzoeken deelwoord (a en b)’.
- (A+B) staat voor ‘de frequentie van alle deelwoorden in de groene [PART+AUX] (A) en rode [AUX+PART] (B) woordvolgorde’.

Op basis van de frequentiegegevens in de tabel wordt aan de hand van dezelfde formule in hoofdstuk 4 voor elk deelwoord de LLR-berekening³⁴ uitgevoerd. Dat levert een score op aan de hand waarvan de statistische significantie bepaald kan worden: hoe hoger de LLR-score, hoe kleiner de p-waarde (i.e. hoe kleiner de kans dat een dergelijk resultaat te wijten is aan het toeval). Een p-waarde die kleiner is dan 0.05 wordt geacht statistisch significant te zijn. In hoofdstuk 4 is daarbij al opgemerkt dat statistische significantie in deze context op twee manieren begrepen kan worden: ofwel komt het deelwoord significant vaak voor in de groene volgorde [PART+AUX] (scenario A) ofwel komt het deelwoord significant niet vaak voor in de groene volgorde [PART+AUX] (scenario B). In dat laatste geval komt het deelwoord significant vaker voor in de rode volgorde [AUX+PART]. Om uit te maken of we met scenario A of met scenario B te doen hebben, maken we gebruik van de volgende vergelijking (de letters staan voor dezelfde frequentiegegevens als in tabel 5.7):

$$(a / A) > (b / B)$$

Het eerste deel van de vergelijking geeft de relatieve proportie weer van het aantal keer dat het deelwoord in de groene volgorde [PART+AUX] voorkomt, het tweede deel geeft de relatieve proportie weer van het aantal keer dat het deelwoord in de rode volgorde [AUX+PART] voorkomt.

Van alle deelwoordtypes in de volledige dataset (dus eindgroepen met een deelwoord en een vervoegde vorm van *zijn*, *hebben* of *worden*) kwamen er 53 (5.81%) statistisch significant vaker voor in de groene volgorde [PART+AUX] en 22 (2.41%) statistisch significant vaker in de rode volgorde [AUX+PART]. In tabel 5.8 hebben we voor de beide volgordes de distinctieve collexemen opgesomd:

Tabel 5.8: Deelwoorden die statistisch significant in de groene [PART+AUX] of rode volgorde [AUX+PART] voorkomen.

Groene volgorde [PART+AUX]	LLR	Rode volgorde [AUX+PART]	LLR
bedoeld	71.82	aangenomen	13.15
bedreigd	45.36	gebracht	12.87
begrepen	22.25	gegaan	11.27
bekend	20.02	gegeven	10.98
beperkt	15.56	gegooid	9.65
bereid	15.56	gehad	9
bestemd	11.1	gekomen	7.23
bestudeerd	8.88	gekregen	6.42
betrokken	8.88	gemaakt	5.81
bevestigd	8.88	genomen	5.62
bevoegd	8.88	gestart	5.62
bewust	6.66	gestemd	5.3
bezorgd	6.66	geweest	5.04
doorzocht	6.66	geworden	4.82
gebaat	6.66	gezet	4.01
gebaseerd	6.66	goedgekeurd	4.01
geblokkeerd	6.66	ontdekt	4.01
gecontroleerd	6.66	ontvangen	4.01
gehecht	6.66	opgeheven	4.01
geïnspireerd	6.66	teruggevonden	4.01
geïnteresseerd	6.49	uitgevoerd	4.01
gekant	6.49	verklaard	3.88
gelezen	6.49		
geneigd	5.93		
geoorloofd	4.67		
gerechtvaardigd	4.67		
gerept	4.67		
gericht	4.44		
geschaad	4.44		
geschikt	4.44		
geslaagd	4.44		
gesteund	4.44		
gevestigd	4.44		
gewonnen	4.44		
gezond	4.44		
onderworpen	4.44		
overleden	4.44		
overspoeld	4.44		
overtuigd	4.44		
toegelaten	4.44		
uitgegroeid	4.44		
uitgesloten	4.44		
verboden	4.44		
verbonden	4.44		
vergemakelijkt	4.44		
verkeerd	4.44		
verplicht	4.44		
verschuldigd	4.44		
vertegenwoordigd	4.44		
vervalst	4.44		
vervuld	4.32		
verwittigd	4.32		
verzekerd	4.26		

Dat er blijkbaar meer deelwoorden zijn die significant vaak in de groene volgorde [PART+AUX] voorkomen (n = 53) dan dat er deelwoorden zijn die significant vaak in de rode volgorde [AUX+PART] voorkomen (n = 22) en, daarmee samenhangend, dat de LLR-waarden voor de significante deelwoorden in de groene volgorde [PART+AUX] gemiddeld genomen een stuk hoger liggen dan de LLR-waarden voor de deelwoorden in de rode volgorde [AUX+PART] (m.n. respectievelijk 8.52 en 6.58; indien men enkel naar de top-vijf van de beide lijsten kijkt, dan is het verschil nog frappanter: respectievelijk 35 en 11.58) wijst erop dat de groene volgorde [PART+AUX] een (lexicaal) specifiekere constructie is dan de rode volgorde [AUX+PART].

Die vaststelling kan in verband gebracht worden met het feit dat de groene volgorde [PART+AUX] in alle omstandigheden gebruikt kan worden, terwijl de rode volgorde [AUX+PART] voornamelijk in aanmerking komt als het deelwoord werkwoordelijk gebruikt is (zie de resultaten van de categorische en de continue analyse van de adjectiviteitsgraad op de woordvolgorde). Dat heeft uiteraard tot gevolg dat deelwoorden die vaak of uitsluitend in adjectivische zin gebruikt worden voornamelijk in de groene volgorde [PART+AUX] verschijnen en bijna nooit in de rode volgorde [AUX+PART], waardoor zo goed als alle adjectivische deelwoorden—voorzover er voldoende frequentie informatie voorhanden is—typerende collexemen zijn voor de groene volgorde [PART+AUX]. Omgekeerd bestaat er geen context waarin enkel de rode volgorde [AUX+PART] mogelijk is, waardoor de rode volgorde [AUX+PART] altijd de concurrentie met de groene volgorde [PART+AUX] moet aangaan en het aantal deelwoorden dat specifiek is voor de rode volgorde [AUX+PART] beperkter is (precies omdat er altijd een tegengewicht is van de groene volgorde [PART+AUX]).

Indien we de typerende deelwoorden in tabel 5.8 van naderbij gaan bestuderen, dan zien we inderdaad dat in de lijst met distinctieve collexemen van de groene volgorde [PART+AUX], i.t.t. de lijst met collexemen voor de rode volgorde [AUX+PART], een aantal deelwoorden opgenomen is dat uitsluitend (bv. *bevoegd*) of voornamelijk (bv. *betrokken*) in een adjectivische zin gebruikt wordt. Die hypothese wordt bevestigd als we kijken naar de adjectiviteitswaarde die aan elk van die deelwoorden toegekend is: in het lijstje met typisch ‘groene’ deelwoorden

[PART+AUX] is aan 10 van de 53 deelwoorden geen adjectiviteitswaarde toegekend, omdat ze niet voorkwamen in combinatie met een vervoegde vorm van *zijn*; in het lijstje ‘rode’ deelwoorden [AUX+PART] is dat 5 van de 22. De gemiddelde adjectiviteitswaarde van de distinctieve collexemen die wel in combinatie met *zijn* voorkwamen, is 0.356 voor de groene volgorde [PART+AUX] en 0.085 voor de rode volgorde [AUX+PART]. Daaruit blijkt dus duidelijk dat de deelwoorden die de groene volgorde [PART+AUX] typeren een heel stuk adjectivischer zijn dan de deelwoorden die de rode volgorde [AUX+PART] typeren.

Indien we tot slot ook de distinctieve collexemen voor de rode volgorde [AUX+PART] van naderbij bekijken, dan zou men kunnen zeggen dat het over het algemeen werkwoorden zijn die makkelijk werkingsgericht geïnterpreteerd kunnen worden—al moet men voorzichtig zijn met zulke uitspraken, omdat de context waarin zo’n deelwoord voorkomt van essentieel belang is voor de interpretatie. Hoe het ook moge zijn, tegen de achtergrond van Pardoens hypothese is de aanwezigheid van *geweest* en *gehad* in de lijst met distinctieve collexemen voor de rode volgorde [AUX+PART] opvallend, aangezien daarvan op geen enkele manier gezegd kan worden dat ze een actionele betekenis hebben. Dat betekent met andere woorden dat de actionele, dynamische betekenis die in het model-Pardoen aan de rode volgorde [AUX+PART] toegekend wordt in sommige omstandigheden in een semantisch conflict kan geraken met de semantiek van het deelwoord. Hierboven is al gesuggereerd dat de taalgebruiker een dergelijk semantisch conflict klaarblijkelijk niet altijd als een probleem ervaart. Aan de hand van de resultaten van de distinctieve-collexemenanalyse kan daaraan toegevoegd worden dat het semantische conflict in sommige omstandigheden zelfs verwaarloosbaar lijkt te zijn, aangezien zelfs deelwoorden met een tegengestelde semantiek de rode volgorde [AUX+PART] kunnen typeren. Die resultaten hoeven geen bewijs contra de semantische hypothese van Pardoen te zijn, maar ze geven wel aan dat bepaalde aspecten van de hypothese op zijn minst aan modificatie toe zijn.

5.2.6 Aanvullingen bij de operationalisering van de categoriale status van het deelwoord

Bij de bespreking van de resultaten in de vorige paragrafen is duidelijk geworden dat de manier waarop we de categoriale status van het deelwoord geoperationaliseerd hebben succesvol blijkt te zijn geweest. Met behulp van het modulair opgebouwde disambigueringsalgoritme (zie paragraaf 5.2.4) hebben we nl. op een objectieve, consistente manier kunnen vaststellen dat de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde in hevige mate beïnvloed wordt door de status van het deelwoord.

Ondanks het feit dat we ervan overtuigd zijn dat het voorgestelde disambigueringsalgoritme een belangrijke stap voorwaarts is voor de studie van een notoir complex fenomeen als dat van de categoriale status van het deelwoord, zal in toekomstig onderzoek toch zeker nagedacht moeten worden hoe het algoritme verder verfijnd en aangevuld kan worden. Zo zal bijvoorbeeld nagegaan moeten worden of het algoritme geoptimaliseerd kan worden door bijvoorbeeld meer rekening te houden met de deelwoordsemantiek: de huidige staat van het algoritme maakt nl. geen onderscheid tussen de verschillende betekenissen van een deelwoord (elk deelwoord krijgt onafhankelijk van het aantal betekenissen waarin het kan voorkomen slechts één waarde) en houdt zelfs op geen enkele manier rechtstreeks rekening met de betekenis waarin een bepaald deelwoord gebruikt is. Dat kan er bijvoorbeeld toe leiden dat een deelwoord dat voornamelijk een actionele, werkingsgerichte betekenis heeft op basis van het algoritme toch als adjectivisch, met een toestandsgerichte betekenis, geclassificeerd wordt.

Verder zal nagegaan moeten worden of een aantal (traditionele) veronderstellingen in het algoritme niet enkel linguïstische constructen zijn, maar ook (cognitieve) realiteitswaarde hebben:

- Module 1 en 2: (semi-)formele deelwoorden kunnen als adjectief beschouwd worden en de geraadpleegde bronnen zijn daarin betrouwbaar (zie paragraaf 5.2.4.2).
- Module 3, submodule A: het intersubjectief gevalideerde criteriumstelsel is representatief (zie paragraaf 5.2.4.2).

- Module 3, submodule B: de ratio van deelwoordvormen en alle werkwoordsvormen zegt iets over de adjectiviteitsgraad; de frequenties in Celex en CGN (R5) zijn representatief; de manier waarop de Celex- en CGN-compileren (R5) de woordsoortinformatie toegevoegd hebben, is accuraat (zie paragraaf 5.2.4.2).

Ook de relatie met de cognitieve strategieën die een taalgebruiker hanteert om in een concrete attestatie de status van het deelwoord te bepalen zal onder de loep genomen moeten worden. Door de modulaire opbouw van het algoritme—detectie van de formele en semi-formele deelwoorden, disambiguering van de ambigue deelwoorden—en door in het algoritme gebruik te maken van contextuele informatie en frequentiegegevens, hebben we er in ieder geval naar gestreefd om, met de beperkte kennis die we hebben, de reële cognitieve disambiguering in zekere mate te imiteren, gebaseerd op de volgende veronderstellingen:

1. *Formele deelwoorden* zijn als aparte adjectivische entiteiten opgeslagen in het mentale lexicon, omdat ze op geen enkele manier meer afgeleid kunnen worden van een basisvorm (de infinitief) en niet meer in verbinding staan met corresponderende werkwoordsvormen.
2. *Semi-formele deelwoorden* kunnen wel nog in verband gebracht worden met een verbaal paradigma, maar omdat ze systematisch divergente semantische of syntactische eigenschappen ontwikkeld hebben, hebben ze zich in zekere mate losgeweekt uit het oorspronkelijk verbale paradigma, waarbij de divergente eigenschappen apart (als adjectief) in het mentale lexicon opgeslagen worden.
3. Deelwoorden die naargelang van de omstandigheden *zowel adjectivisch als verbaal* geïnterpreteerd kunnen worden, worden in de eerste plaats gedisambigüeerd op basis van elementen die contextueel voorhanden zijn. Indien de context geen uitsluitel biedt, dan kan de taalgebruiker aan de hand van stochastische informatie die hij/zij experiëntieel vergaard heeft het deelwoord trachten te disambigueren, met name in die zin dat de taalgebruiker bij ontstentenis van contextuele indicatoren voor de frequentste interpretatie kiest—waarbij frequentie correleert met cognitieve toegankelijkheid en prominentie. Een dergelijke zienswijze impliceert dat de frequentie waarmee een deelwoord in een bepaalde betekenis gebruikt wordt, bepaalt hoe dat

deelwoord in ons geheugen opgeslagen is en dat, met andere woorden, het taalsysteem niet vast en onveranderlijk is, maar constant adapteert aan het taalgebruik (dat op zijn beurt uiteraard vorm krijgt door het taalsysteem). Een dergelijke visie op de disambiguering van het deelwoord past perfect binnen wat we in hoofdstuk 1 aangekondigd hebben als een emergente benadering van taal, i.e. het opheffen van de grens tussen taalsysteem en taalgebruik en een voortdurende en dynamische wisselwerking tussen taalsysteem en taalgebruik. Ondanks het feit dat de cognitieve hypothesen over de disambiguering van ambigue deelwoorden nog volledig getest moet worden in een psycho-experimenteel design, kan men op basis van reeds beschikbare onderzoeksresultaten in verwante gebieden toch al zeggen dat er inderdaad een duidelijke invloed lijkt te zijn van contextuele informatie en frequentie op de manier waarop talige kennis in ons geheugen gestructureerd is, beschikbaar en toegankelijk is—“Frequency [...] affects, first of all, cognitive processes” (Fenk-Oczlon 2001: 232)—en bijdraagt aan de disambiguering van lexicale elementen (zie bv. Simpson 1994).

5.3 De impact van de aard van het hulpwerkwoord

In de vorige sectie hebben we de invloed van de categoriale status van het deelwoord op de woordvolgorde bestudeerd voor werkwoordelijke eindgroepen die opgebouwd zijn uit een deelwoord en een vervoegde vorm van *zijn*. In deze sectie verbreden we het perspectief, door naast eindgroepen met *zijn* ook eindgroepen met de hulpwerkwoorden *hebben* en *worden* in de analyse te betrekken. We beginnen met de bespreking van de belangrijkste literatuur ter zake (paragraaf 5.3.1) en de hypothesen die in deze sectie getest zullen worden (paragraaf 5.3.2). In paragraaf 5.3.3 worden vervolgens de resultaten van de analyses voorgesteld en besproken.

5.3.1 *Status quaestionis*

Naast de status van het deelwoord is het onderzoek naar de semantische dimensie van de woordvolgordevariatie ook bestudeerd aan de hand van het type hulpwerkwoord waarmee de deelwoorden gecombineerd kunnen worden. In tegenstelling tot de studies

die zich uitsluitend op de status van het deelwoord gefocust hebben, is het onderzoek naar de aard van het hulpwerkwoord grotendeels empirisch-kwantitatief van aard. Dat is niet verwonderlijk, aangezien de factor makkelijker te operationaliseren is. Toch heeft ook de factor *aard van het hulpwerkwoord* te lijden onder problemen, zij het deze keer van interpretatieve aard: hoe moeten de verschillen tussen de types hulpwerkwoorden precies begrepen worden? We geven een kort overzicht van de belangrijkste bevindingen van dat onderzoek (voor een schematisch overzicht: zie het einde van deze paragraaf).

Pauwels (1953: 115-118) maakt als eerste een onderscheid naar de aard van het hulpwerkwoord en komt daarbij tot de vaststelling dat eindgroepen met *zijn* minder frequent in de groene volgorde [PART+AUX] voorkomen (72.7%) dan *hebben* (75.4%) en *worden* (78%).³⁵ Tegen de achtergrond van Pardoens hypothese (paragraaf 5.2.2) is dat een opmerkelijk resultaat: vanuit de traditionele redenering dat *zijn* als koppelwerkwoord gebruikt kan worden, het deelwoord daardoor adjectivisch is en het adjectivische karakter van het deelwoord (en de toestandsinterpretatie van de zin als geheel) de groene volgorde [PART+AUX] bevordert (volgens Pardoens is dat zelfs de enige mogelijkheid), zouden we nl. een grotere proportie groene volgordes [PART+AUX] verwachten. Een mogelijke reden waarom de proportie groene volgordes [PART+AUX] veeleer bescheiden is, kan wellicht gevonden worden in de aard van de aangeboden items. Als we nl. gaan kijken welke items Pauwels aan haar informanten aangeboden heeft, dan stellen we vast dat geen enkele van de deelwoorden die in combinatie met *zijn* voorkomen in aanmerking lijkt te komen voor een adjectivische interpretatie. Verder moet ook nog opgemerkt worden dat Pauwels de resultaten niet differentieert naar verschillende regio's, terwijl bijvoorbeeld voor het Oost-Vlaamse materiaal vastgesteld kan worden dat het meer groene volgordes [PART+AUX] in de zinnen met *zijn* bevat dan in zinnen met *hebben* en *worden*. Naast het onderzoek naar het effect van het hulpwerkwoord in het materiaal dat ze op basis van enquêtes verzameld heeft, bestudeert Pauwels (1953: 262-263) ook nog in literair proza de invloed van het hulpwerkwoord en komt daarbij tot de vaststelling dat clusters met *worden* het vaakst de groene volgorde [PART+AUX] vertonen (63.0%, n= 181), gevolgd door *zijn* (62.0%, n = 476) en *hebben* (57.4%, n = 698). Wat maakt dat de posities van *zijn* en *hebben* ten opzichte van het dialectologische materiaal

omgekeerd zijn, is niet geheel duidelijk. In ieder geval zijn de verschillen tussen de types hulpwerkwoorden in het literaire proza niet significant (eigen analyse).

De Schutter (1964: 68-69) vindt in zijn corpus met literaire teksten (n = 2284) min of meer dezelfde tendensen weer: ook bij hem komen het minst groene volgordes [PART+AUX] voor met het hulpwerkwoord *hebben* (52.7%), gevolgd door 55.2% in eindgroepen met actief *zijn*, 62.1% met *worden* en 69.1% groene volgordes [PART+AUX] in werkwoordelijke eindgroepen met passief *zijn*. De verschillen zijn volgens eigen berekeningen significant ($\chi^2 = 29.20$, $df = 3$, $p < .0001$), maar bij nadere beschouwing blijken op basis van een gepartitioneerde chi-kwadraatanalyse de verschillen tussen de eindgroepen met *hebben* en actief *zijn* en tussen eindgroepen met passief *worden* en *zijn* niet statistisch significant te zijn. Dat eindgroepen met passief *zijn* een voorliefde lijken te hebben voor de groene volgorde [PART+AUX] brengt De Schutter in verband met een analogiewerking vanuit de constructie adjectief + koppelwerkwoord, waar geen of nauwelijks variatie is, maar hij merkt daarnaast toch op dat in de categorie met passief *zijn* ook een aantal twijfelgevallen opgenomen zijn van zinnen met *zijn* die zowel als hulpwerkwoord van het passief of als koppelwerkwoord beschouwd kunnen worden, met een groter aandeel groene volgordes [PART+AUX] in die categorie als gevolg. De Schutter gewaagt verder ook van een dergelijk analogie-effect op de constructie met een hulpwerkwoord van tijd (die er formeel op lijkt); dat zou verklaren waarom de eindgroepen met *hebben* de kleinste proportie groene volgordes [PART+AUX] hebben. Waarom de groep met *worden* als hulpwerkwoord dan een zo groot aandeel groene volgordes [PART+AUX] heeft, vermeldt De Schutter niet.³⁶

De resultaten van De Schutters 1964-studie wordt in de daaropvolgende decennia door verschillende onderzoekers geverifieerd, meer bepaald door De Schutter (1976), De Cubber (1986), Haeseryn (1990), De Schutter (1996) en Swerts (1998). We bespreken kort de resultaten van die onderzoeken.

In 1976 doet De Schutter (1976: 232) zijn analyse van 1964 nog eens over, maar hij behoudt slechts twee categorieën, m.n. passief en actief. Hij constateert dezelfde tendensen als in de studie van 1964: in actieve bijzinnen vindt hij 48.1% (n = 649) van de eindgroepen in groene volgorde [PART+AUX], in passieve bijzinnen 61.3% (n = 326); het verschil tussen de beide groepen is significant ($\chi^2 = 13.82$, $df = 1$, $p < .0001$;

eigen berekeningen), maar het is niet geheel duidelijk hoe dat resultaat geïnterpreteerd moet worden.³⁷

De Cubber (1986: 66-67) bestudeert het effect van het hulpwerkwoord aan de hand van dagboekmateriaal en constateert daarbij dat in 84.2% van de eindgroepen met passief *zijn* de groene volgorde [PART+AUX] gekozen wordt, 79.8% met actief *zijn*, 77.5% met *worden* en 64.2% met *hebben*. Dat verschil is opnieuw significant (eigen berekeningen: $\chi^2 = 14.19$, $df = 3$, $p < .01$). Een gepartitioneerde chi-kwadraatanalyse laat echter zien dat de vier groepen teruggebracht kunnen worden naar twee groepen, met name de groep met *hebben* versus de rest. De Cubber interpreteert de resultaten op dezelfde manier als De Schutter (1964), ondanks het feit dat de relatieve positie van de categorie met actief *zijn* van plaats gewisseld heeft met de categorie *worden*.

Haeseryn (1990: 144-146) vindt aan de hand van zijn experimentele items (A408), (B203), (B211) en (B113) dat eindgroepen met passief *zijn* in 74.40% van de gevallen de groene volgorde [PART+AUX] vertonen, terwijl dat voor eindgroepen met *hebben* slechts 57.44% is ($\chi^2 = 32.49$, $df = 1$, $p < .0001$).³⁸

In het gesproken materiaal van De Schutter (1996: 214, 220) worden dezelfde tendensen als De Schutter 1976 bevestigd: 70.11% groene volgordes [PART+AUX] in eindgroepen met een hulpwerkwoord van het passief en 52.40% met een ergatief hulpwerkwoord ($n = 403$; $\chi^2 = 12.76$, $df = 1$, $p < .001$).

Ook Swerts (1998: 302-303) onderzoekt het effect van het type hulpwerkwoord in gesproken taal. Hij merkt daarbij op dat eindgroepen met *zijn* (vooral passief) 60.71% groene volgordes [PART+AUX] vertonen, gevolgd door *hebben* met 30% en *worden* met 20% groene volgordes [PART+AUX]. Opmerkelijk hier is de positie van *worden* die in de andere studies veeleer met de groene volgorde [PART+AUX] verbonden werd. Hoe dat verschil precies geïnterpreteerd moet worden, is opnieuw niet heel erg duidelijk, maar in ieder geval moet er wel rekening mee worden gehouden dat de resultaten van Swerts het best met enige voorzichtigheid bejegend worden, aangezien zijn onderzoek op weinig materiaal gebaseerd is ($n = 71$; 50% van de cellen heeft een frequentie van minder dan 5) en het daarom onzeker is of zijn cijfers representatief zijn voor de volledige populatie.

In wat volgt zullen we voor onze eigen dataset nagaan in hoeverre die resultaten van vroeger onderzoek bevestigd kunnen worden en in welke mate de voorgestelde interpretaties behouden kunnen blijven. Vooraleer we daaraan beginnen vatten we de resultaten van het vroegere onderzoek kort nog even samen aan de hand van tabel 5.9 die voor elk artikel/monografie de relatieve voorkeur voor de groene [PART+AUX] of rode volgorde [AUX+PART] uitsplitst over de aard van het hulpwerkwoord.

Tabel 5.9: Relatieve positie van de verschillende hulpwerkwoorden volgens de proportie groene woordvolgordes [PART+AUX] in vroeger onderzoek.

Auteur	Proportie groene woordvolgordes [PART+AUX]
Pauwels (1953)	Enquête: worden > hebben > zijn Proza: worden > zijn > hebben
De Schutter (1964)	Zijn (passief) > worden > zijn (actief) > hebben
De Schutter (1976)	Passief > actief
De Cubber (1986)	Zijn (passief) > zijn (actief) > worden > hebben
Haeseryn (1990)	Zijn (passief) > hebben
De Schutter (1996)	Passief > ergatief
Swerts (1998)	Zijn > hebben > worden

5.3.2 Hypotheses

Uit tabel 5.9 is duidelijk gebleken dat de bevindingen van eerder onderzoek nogal sterk uiteenlopen, waardoor het niet eenvoudig is om er bruikbare hypothesen aan te onttrekken. We hebben ervoor geopteerd om [HYP 5.3] te baseren op sterke tendensen binnen het vroegere onderzoek. [HYP 5.4] is identiek aan de in paragraaf 5.3.1 voorgestelde analogiehypothese van De Schutter (1964).

[HYP 5.3] Werkwoordelijke eindgroepen die het hulpwerkwoord *zijn* bevatten, komen vaker in de groene volgorde [PART+AUX] voor dan eindgroepen met het hulpwerkwoord *hebben* of *worden* (paragraaf 5.3.3).

[HYP 5.4] Passieve constructies komen vaker in de groene volgorde [PART+AUX] voor dan actieve, niet-copulatieve constructies (paragraaf 5.3.3).

5.3.3 Resultaten en discussie

In navolging van Pauwels (1953) en Swerts (1998) hebben we de eindgroepen in onze dataset in een eerste stap opgesplitst naar de vorm die het hulpwerkwoord (of koppelwerkwoord in het geval van een eindgroep met *zijn* en een adjectivisch deelwoord) aanneemt. De hypothese is de volgende:

[HYP 5.3] Werkwoordelijke eindgroepen die het hulpwerkwoord *zijn* bevatten, komen vaker in de groene volgorde [PART+AUX] voor dan eindgroepen met het hulpwerkwoord *hebben* of *worden*.

Tabel 5.10: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type hulpwerkwoord.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
<i>Zijn</i>	46.91% (395/841)	53.09% (446/841)
<i>Hebben</i>	18.98% (127/665)	81.02% (538/665)
<i>Worden</i>	30.32% (268/884)	69.68% (616/884)

Tabel 5.10 laat zien dat werkwoordelijke eindgroepen die bestaan uit een deelwoord en een vervoegde vorm van *zijn* de hoogste proportie groene woordvolgordes [PART+AUX] vertonen (bevestiging [HYP 5.3]), gevolgd door eindgroepen met

worden en *hebben*, al moet daarbij onmiddellijk opgemerkt worden dat de groene volgorde [PART+AUX] nergens de dominante volgorde is (het relatieve aandeel groene volgordes [PART+AUX] stijgt niet boven de 46.91% uit). Die relatieve volgorde komt min of meer overeen met de volgorde die De Cubber (1986) vond op basis van zijn dagboekmateriaal. Het verschil in verdeling tussen de types hulpwerkwoorden is in ieder geval statistisch significant ($\chi^2 = 135.10$, $df = 2$, $p < .0001$; $\lambda = 0.06$, $ASE = 0.01$), wat betekent dat we voor meer dan 99% zeker kunnen zijn dat er een echt verband is tussen de keuze van de woordvolgorde en het hulpwerkwoord.

Dat eindgroepen met *zijn* de hoogste proportie groene woordvolgordes [PART+AUX] vertoont, hoeft niet te verwonderen tegen de achtergrond van de resultaten van vroeger onderzoek en de bovenstaande analyses (paragraaf 5.2.5). De groep met *zijn* bevat nl. alle gevallen waarin *zijn* als koppelwerkwoord (gecombineerd met een adjectivisch deelwoord) gebruikt is—waarvan we in paragraaf 5.2.5 vastgesteld hebben dat ze een sterke voorkeur voor de groene volgorde [PART+AUX] hebben.

Daarnaast is niet uitgesloten, zoals De Schutter (1964 en 1996) suggereert, dat de andere eindgroepen met niet-copulatief (passief) *zijn* door een analogie-effect ook vaker in de groene volgorde [PART+AUX] verschijnen: “Het is nogal evident dat passieve constructies in de meeste gevallen de volle nadruk leggen op het resultaat van de handeling en/of proces; daardoor krijgt het VD [voltooid deelwoord, GDS] ten minste gedeeltelijk een ‘adjectivische’ connotatie [...]” (De Schutter (1996: 217-218) en wordt het conceptueel aangetrokken tot de constructie adjectivisch deelwoord+koppelwerkwoord.

[HYP 5.4] Passieve constructies komen vaker in de groene volgorde [PART+AUX] voor dan actieve, niet-copulatieve constructies.

Om [HYP 5.4] te testen, zullen we, in een tweede stap, in tabel 5.11, gebruik makend van de categorische classificatie van deelwoorden die gecombineerd worden met een vervoegde vorm van *zijn* (paragraaf 5.2.4.2 en 5.2.5.1), de groep met *zijn* verder opsplitsen naar de functie: koppelwerkwoord (uit de gepartitioneerde chi-kwadraatanalyse hierboven was gebleken dat er geen verschillen zijn tussen de

verschillende types adjectivische deelwoorden/koppelwerkwoorden), hulpwerkwoord van het passief en hulpwerkwoord van tijd.

Tabel 5.11: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type hulpwerkwoord.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Copulatief <i>zijn</i>	86.82% (188/215)	13.18% (27/215)
Hulpwerkwoord van tijd: <i>zijn/hebben</i>	20.09% (191/942)	79.91% (751/942)
Passief <i>zijn</i>	37.05% (103/278)	62.95% (175/278)
<i>Worden</i>	30.23% (268/884)	69.77% (616/884)
Onclassificeerbaar <i>zijn</i>	56.34% (40/71)	43.66% (31/71)

Op basis van een gepartitioneerde chi-kwadraatanalyse zijn we te weten gekomen dat er geen statistisch significant verschil is tussen de hulpwerkwoorden van tijd *zijn* en *hebben*, zodat we ze voor de statistische analyse samengenomen hebben. Die statistische analyse laat zien dat het verschil in verdeling tussen de types hulpwerkwoorden opnieuw statistisch significant is ($\chi^2 = 379.30$, $df = 4$, $p < .0001$; $\lambda = 0.11$, $ASE = 0.01$). Belangrijker is echter dat de analogiehypothese van De Schutter (1996) grotendeels bevestigd kan worden (bevestiging van [HYP 5.4]): de passieve constructies verschijnen statistisch significant vaker in de groene volgorde [PART+AUX], zij het niet in dezelfde mate als de eindgroepen met een koppelwerkwoord, waardoor ze een tussenpositie innemen tussen de koppelwerkwoorden en de constructies met een hulpwerkwoord van tijd—dat eindgroepen met *zijn* als koppelwerkwoord het vaakst in de groene volgorde [PART+AUX] voorkomen, hoeft uiteraard niet te verwonderen, aangezien de indeling gebaseerd is op het disambigueringsalgoritme en hierboven al vastgesteld is dat clusters adjectivisch deelwoord+koppelwerkwoord zeer vaak in de groene volgorde [PART+AUX] voorkomen.

De Schutter (1964) suggereerde verder nog dat een dergelijk analogie-effect ook van toepassing zou kunnen zijn op de hulpwerkwoorden van tijd, maar aan de hand van onze data vinden we daar geen bevestiging voor: eindgroepen met een hulpwerkwoord van tijd komen veel vaker in de rode volgorde [AUX+PART] voor.

Uiteraard kunnen we de vraag naar een verklaring van de verschillen tussen de verschillende types hulpwerkwoorden ook omkeren, bijvoorbeeld door ons af te vragen waarom eindgroepen met een hulpwerkwoord van tijd zo vaak in de rode volgorde [AUX+PART] voorkomen. Het antwoord op die vraag ligt veel minder voor de hand en ook in de literatuur werden totnogtoe geen suggesties gedaan om die voorkeur te verklaren. Een mogelijke verklaring zou kunnen zijn dat, verder redenerend op de argumentatie van De Schutter (1964 en 1996), aan eindgroepen met een hulpwerkwoord van tijd, anders dan aan eindgroepen met een koppelwerkwoord of met een passief hulpwerkwoord, conceptueel minder snel een adjectivische connotatie toegekend kan worden, i.e. een toestandssemantiek, zodat eindgroepen met een hulpwerkwoord van tijd de ideale omgeving zijn om de sociostilistische (hoofdstuk 3), prosodische (hoofdstuk 4) of psycholinguïstische (hoofdstuk 6) mechanismen ten volle te laten werken. Het spreekt voor zich dat een dergelijke tentatieve verklaring verder onderzocht zal moeten worden.

5.4 Samenvatting

In dit hoofdstuk hebben we de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep met een deelwoord en een hulpwerkwoord vanuit een semantisch perspectief bestudeerd. De analyses waarover in dit hoofdstuk gerapporteerd is, hebben daarbij duidelijk gemaakt dat de keuze voor de groene [PART+AUX] of rode [AUX+PART] volgorde in hevige mate bepaald wordt door de twee intergerelateerde factoren *categoriale status van het deelwoord* (is het deelwoord werkwoordelijk of adjectivisch gebruikt?) en *aard van het hulpwerkwoord* (welk type hulpwerkwoord wordt met het deelwoord gecombineerd?). We vatten voor de beide factoren de belangrijkste resultaten samen.

5.4.1 *Categoriale status van het deelwoord*

In een eerste stap zijn we nagegaan of er empirische evidentie gevonden kan worden voor de traditionele veronderstelling dat een adjectivisch gebruikt deelwoord (i.t.t. een werkwoordelijk gebruikt deelwoord) enkel in de groene volgorde [PART+AUX] kan voorkomen. Daartoe zijn we eerst ruim blijven stilstaan bij de vraag hoe een adjectivisch deelwoord van een werkwoordelijk deelwoord op een objectieve en systematische manier onderscheiden kan worden. Aangezien de criteria die de traditionele taalkunde biedt als ontoereikend beschouwd werden, hebben we zelf een disambigueringsalgoritme opgesteld waarmee op een objectieve en eenduidige manier vastgesteld kan worden of en in welke mate een deelwoord (gecombineerd met een vervoegde vorm van het werkwoord *zijn*) adjectivisch of verbaal gebruikt is. Het algoritme bestaat uit drie modules en maakt zowel gebruik van contextgevoelige als van contextongevoelige indicatoren (zoals frequentie-informatie) om de categoriale status van het deelwoord te bepalen.

De resultaten van de analyses (enkel voor de eindgroepen met een vervoegde vorm van *zijn*) laten zien dat de keuze voor de woordvolgorde inderdaad op een overweldigende manier bepaald lijkt te worden door de status van het deelwoord: een deelwoord dat op een adjectivische manier gebruikt wordt, verschijnt vaker in de groene volgorde [PART+AUX] dan een deelwoord dat op een werkwoordelijke manier gebruikt wordt. Echter, in tegenstelling tot wat traditioneel aangenomen wordt, geven de resultaten ook aan dat er op geen enkele manier sprake kan zijn van een categorisch effect van de status van het deelwoord, i.e. de adjectivische status van het deelwoord is niet per se incompatibel met de rode volgorde [AUX+PART]. Uit de gegevens in onze dataset blijkt nl. dat een niet onaanzienlijke fractie adjectivische deelwoorden in de rode volgorde [AUX+PART] verschijnt. Bij een detailanalyse van de groep met werkwoordelijke deelwoorden bleek tot slot nog dat de passieve deelwoorden een middenpositie bekleden tussen de adjectivische deelwoorden en de voltooide deelwoorden wat het gebruik van de groene woordvolgorde [PART+AUX] betreft.

In een tweede stap is op exploratieve wijze, aan de hand van een zgn. *distinctieve-collexemenanalyse*, nagegaan of de hypothese-Pardoen bevestigd kan

worden. De hypothese-Pardoen zegt dat bij een eindgroep in de rode volgorde de gedachte aan een werking allesoverheersend is. Indien men op basis van die hypothese aanneemt dat door de aard van de constructionele semantiek van de rode volgorde (dynamische, actionele betekenis) een toestandsgericht deelwoord niet in de rode volgorde kan verschijnen (hoe moet men een toestandsbetekenis rijmen met een dynamische betekenis?), dan maakt de distinctieve-collexemenanalyse in ieder geval duidelijk dat taalgebruikers in sommige omstandigheden geen graten zien in een semantisch conflict tussen de constructionele en participiale semantiek. Uit de lijst met deelwoorden die typisch in de rode volgorde voorkomen, staat bijvoorbeeld een deelwoord als *geweest*. Die vaststelling hoeft uiteraard niet te betekenen dat de hypothese-Pardoen volledig gefalsificeerd is, maar er blijkt wel uit dat de hypothese in zekere mate genuanceerd moet worden.

5.4.2 Aard van het hulpwerkwoord

In het tweede deel van het hoofdstuk werd het perspectief verbreed, door naast eindgroepen met een vervoegde vorm van *zijn* ook andere types hulpwerkwoorden (i.c. *hebben* en *worden*) in de analyse te betrekken. Zoals verwacht kon worden op basis van de resultaten die in paragraaf 5.4.1 voorgesteld zijn, geven de analyses enerzijds aan dat werkwoordelijke eindgroepen met een vervoegde vorm van *zijn* vaker in de groene volgorde [PART+AUX] voorkomen, gevolgd door respectievelijk *worden* en *hebben*. Anderzijds kon vastgesteld worden dat de eindgroepen met een hulpwerkwoord van het passief vaker in de groene volgorde [PART+AUX] voorkomen dan eindgroepen met een hulpwerkwoord van tijd. Die laatste vaststelling werd in verband gebracht met de analogiehypothese van De Schutter (1964), i.e. doordat in een passieve constructie de nadruk gelegd wordt op het resultaat van de handeling of het proces, krijgt het deelwoord ten minste gedeeltelijk een adjectivische connotatie, waardoor de passieve constructie conceptueel aangetrokken wordt door de constructie met een adjectivisch deelwoord (die, zoals bekend, erg frequent in de groene volgorde [PART+AUX] voorkomt).

Noten:

1. In de rest van dit hoofdstuk wordt de term *deelwoord* gereserveerd voor de formele categorie zonder semantische of functionele implicaties. Indien een onderscheid gemaakt moet worden in verschillende types, semantisch onderscheiden, deelwoorden, dan wordt er een extra toevoeging bij geplaatst: adjectivisch deelwoord versus verbaal of werkwoordelijk deelwoord.
2. Een dergelijke woordklasseverschuiving is geen uniek fenomeen: “The categories we work with in the real world are not homogeneous [...], and word classes are no exception. [...] Just as the characteristics of the members of a class may vary, there may be unclear borderlines between the characteristics of the members of one class and another [...]. We set up grammatical classes, as well as semantic classes and syntactic roles, because we note that elements show similarities in their behaviour. But the flexibility and complexity of language defy our neat classification” (Biber et al. 1999: 59). Wanneer deelwoorden ook adjectivische eigenschappen zijn beginnen vertonen en hoe die ontwikkeling zich verhoudt ten opzichte van de ontwikkeling van het deelwoord uit de predicatieve bepaling, is niet geheel duidelijk.
3. Niet iedereen gaat akkoord met de zienswijze dat het deelwoord een ambivalente semantiek heeft. Voor Pardoën (1991) en Verhagen (1992) duidt het deelwoord een atemporele toestand aan die begrepen moet worden als een toestand ten gevolge van een werking (resultatief, maar niet terminatief). In die toestandsaanduiding wordt volledig geabstraheerd van temporele verwijzingen, zodat noch de toestand, noch de werking speciale aandacht krijgen. Het is door het gebruik van een deelwoord in een specifieke context dat de gedachte aan de werking dan wel de gedachte aan een toestand op de voorgrond treedt.
4. Algemeen wordt aangenomen dat een duratieve bepaling een adjectivische interpretatie in de hand werkt. We sluiten ons daar voorlopig bij aan. In het verdere verloop van dit hoofdstuk zullen we die traditionele assumptie kritisch evalueren.
5. Het is niet geheel duidelijk welke relatie tussen de status van het finiete werkwoord en het deelwoord verondersteld mag worden: (i) de interpretatie van het finiete werkwoord bepaalt hoe het deelwoord geïnterpreteerd moet worden; (ii) de interpretatie van het deelwoord bepaalt hoe het finiete werkwoord geïnterpreteerd moet worden; (iii) de interpretatieverschuiving vindt gelijktijdig plaats.
6. Over de vraag of we in het geval van een historisch deelwoord als *bereid* mogen spreken van een pseudodeelwoord lijken de meningen verdeeld: de ANS (1997: 725) beperkt de term *pseudodeelwoord* tot desubstantivische afleidingen van het type *gerokt*, terwijl bijvoorbeeld Smedts & Van Belle (1994: 252) wel historische deelwoorden als pseudodeelwoord beschouwen. In ieder geval is het zo dat vraagtekens geplaatst kunnen worden bij de onmiddellijke incorporatie van pseudodeelwoorden, en vooral dan historische deelwoorden, in de categorie der adjectieven: deelwoorden als *geboren*, bijvoorbeeld, dragen immers een onmiskenbaar werkwoordelijk aspect dat onmogelijk compatibel kan zijn met de algemene semantische karakteristieken van een adjectief. Het lijkt er dan ook op dat dergelijke deelwoorden voornamelijk op grond van het gebrek aan een volledig werkwoordelijk paradigma als adjectief beschouwd worden, en niet zozeer op

semantische gronden. In het vervolg van deze tekst zullen we de term pseudodeelwoord dan ook gebruiken om te verwijzen naar de klasse van formeel op een deelwoord lijkende woorden zonder volledig verbaal paradigma.

7. Het redeneerschema dat Michels hanteert, ziet er als volgt uit: een feit (sommige deelwoorden verschijnen zeer frequent in de groene volgorde [PART+AUX]) roept een verklaring op (die deelwoorden zijn adjectivisch) en die verklaring wordt op haar beurt gebruikt om een absolute, normatieve uitspraak te doen (als een deelwoord adjectivisch is, mag je het niet in de rode volgorde [AUX+PART] gebruiken of als je een rode volgorde [AUX+PART] gebruikt dan is de interpretatie verplicht verbaal).
8. Pardoens lijkt zelf haar analyse niet te beschouwen als een verklaring voor het interpretatieverschil tussen de groene [PART+AUX] en de rode volgorde [AUX+PART]. De analyse die ze biedt kan echter wel zo opgevat worden.
9. Van der Horst gaat in zijn studie, die hij kadert binnen de analytische taalkunde, akkoord met Pardoens visie, maar formuleert het wat anders. Kort samengevat komt het hier op neer: het hoofdwerkwoord is de semantische en syntactische kern van de zin; bij het passeren van het hoofdwerkwoord kan de interpretatie van de zin (voorlopig) afgerond worden. Voor de groene volgorde [PART+AUX] betekent dat dan dat de zin geïnterpreteerd kan worden zonder het finiete werkwoord, voor de rode volgorde [AUX+PART] met het finiete werkwoord. Gegeven de soort betekenis van deelwoorden en hulpwerkwoorden (zie Pardoens) is dat dan voldoende om het onderscheid tussen statisch en dynamisch te verklaren.

Duinhoven (1998) borduurt hierop verder en tracht aan te geven waarom het deelwoord in de groene volgorde [PART+AUX] liever verbonden wordt met het preverbale materiaal dan met het finiete werkwoord. Hij brengt daarvoor de volgordevariatie in verband met het onderscheid tussen synthetische en analytische constructies. Een analytische constructie onderscheidt zich van een synthetische doordat ze de elementen van de constructie apart, het een na het ander, uiteenzet, terwijl een synthetische constructie de constituerende elementen als een geheel presenteert. Typerend voor een analytische constructie is dat de kern voorop staat, terwijl dat bij een synthetische constructie niet per se het geval hoeft te zijn (dus: zowel de volgorde kern-bepaling als bepaling-kern zijn daar mogelijk). Duinhoven betoogt verder dat de werkwoordelijke eindgroep in de loop der tijden gesynthetiseerd is. Als de werkwoordelijke eindgroep nu in de groene volgorde voorkomt [PART+AUX], i.e. eerst de kern (deelwoord) en dan de bepaling (finiet werkwoord), dan is dat uiteraard perfect compatibel met de synthetische constructie, maar de omgekeerde volgorde (bepaling-kern) zou de synthetische aard van de constructie onmiskenbaar duidelijk maken. Daardoor (en door het feit dat het finiete werkwoord de kern is van de bijzin) zou de suggestie gewekt kunnen worden, zo meent Duinhoven, dat het finiete werkwoord in de groene volgorde [PART+AUX] als de kern van de werkwoordsgroep beschouwd dient te worden en het deelwoord een ondergeschikte status heeft. Daar het deelwoord niet ondergeschikt geacht kan worden aan het finiete werkwoord, wordt het met het preverbale materiaal verbonden.

10. Het is niet ondenkbaar dat bij de voorstelling van de semantische hypothese Pardoens de redeneerfout heeft gemaakt dat de rode volgorde [AUX+PART] onlosmakelijk verbonden is met

een werkwoordelijk gezegde (wat al gecontesteerd kan worden) en dat een werkwoordelijk gezegde werkwoordelijk, i.e. werkingsgericht, geïnterpreteerd moet worden. Haar voorbeelden zijn daarbij verhelderend en intuïtief volstrekt aanvaardbaar, maar ze maakt daarbij de fout een deel van de deelwoorden in de rode volgorde [AUX+PART] die een minder prototypische actionele interpretatie hebben, links te laten liggen.

11. Ondanks wat hierboven opgemerkt is over de band tussen adjectivische deelwoorden en naamwoordelijke gezegdes volgen we voorlopig de traditionele zienswijze dat enkel deelwoorden die gecombineerd worden met het werkwoord *zijn* als potentieel ambigu beschouwd mogen worden. Toekomstig onderzoek zal moeten nagaan in hoeverre het hier te ontwikkelen algoritme ook op andere werkwoorden (zoals deelwoorden gecombineerd met *hebben*) toegepast moet worden.
12. We zullen ervan uitgaan dat pseudodeelwoorden tot de klasse der adjectieven gerekend moeten worden. Het valt echter te betwijfelen of dat voor alle pseudodeelwoorden volgehouden kan worden. Zo lijkt het er bijvoorbeeld op dat historische deelwoorden traditioneel als adjectief beschouwd worden, omdat er geen werkwoordelijk paradigma meer is (een formele reden) en niet om semantische redenen (er is sprake van een toestandsgerichte semantiek). *Geboren* is daar een mooi voorbeeld van: traditioneel wordt het als adjectief beschouwd, maar toch kan het perfect met een bijwoordelijke bepaling van plaats en tijd gecombineerd worden, net als werkwoordelijk gebruikte en werkingsgerichte deelwoorden: *Hij is op 14 juni 1977 in Aalst geboren*.
13. Ook wel participia praeverbalia genoemd: “Andere adjectieven met ge- gelijken wat hun vorming betreft op deelwoorden, maar zijn van substantieven afgeleid, als gespierd, gelaarsd, gebloemd, gespikkeld, geharnast, geletterd, genaamd, met de betekenis: voorzien van wat het grondwoord aanduidt. Zulke afleidingen, door Van Haeringen participia praeverbalia genoemd, bieden ‘een praktisch middel om een meer omslachtige constructie met voorzetsels of voegwoorden te vermijden” (De Vooyoys & Schönfeld 1957: 239-240).
14. De 20 deelwoorden waarvan het vermoeden bestond dat ze als formeel beschouwd dienden te worden, maar waarvan achteraf bleek dat dat toch niet het geval was, zijn: *aangeslagen, aangewezen, bekend, besteed, bezorgd, gedaan, geïnteresseerd, geleerd, genegen, geschikt, gespecificeerd, gesteld, ingenomen, opgewassen, toegedaan, uitgesproken, verdeeld, verkeerd, vertrouwd, vooringenomen*.
15. Het is onduidelijk of bij de ontwikkeling van deelwoorden naar adjectieven een lexicalisatieproces verondersteld mag worden dan wel een semasiologisch extentie- of reductieproces. In de Celex-documentatie (Baayen et al. 1993) wordt in ieder geval voor lexicalisatie gekozen: “Typically, present and past participles become independent adjectives. In the *Woordenlijst van de Nederlandse Taal*, the word *begrensd* is listed as a bold type entry in its own right as well as a flexion of the verb *begrenzen*. Forms such as these are called lexicalised flexions”.
16. We maken hier geen gebruik van de semantische criteria, omdat daarvan vastgesteld was dat ze moeilijk objectiveerbaar en operationaliseerbaar zijn. De morfologische tests zijn op hun beurt ook niet bruikbaar om de adjectiefrealiteit vast te stellen, omdat ze enkel uitsluitel geven over de

mogelijkheid om het deelwoord in een adjectivische context (met een adjectivische interpretatie) te plaatsen. Een uitzondering daarop was het morfologische criterium dat gebaseerd was op de woordsoort van de verbale stam: indien die stam een nomen was, dan moest het deelwoord als een adjectief worden beschouwd. Aangezien dergelijke desubstantivische afleidingen (participia praeverbalia) niet ingepast kunnen worden in een werkwoordelijk paradigma, zijn die in de eerste module al gedetecteerd en heeft het daarom geen zin om het criterium hier te herhalen.

17. De informanten waren Nele Nivelles, Karen Deschamps, Geert Brône, Jose Tummers en Sofie Van Gijssel, allen werkzaam aan het departement Linguïstiek van de K.U.Leuven. Geen van de onderzoekers had een meer dan gemiddelde kennis over het onderzoek naar de participiale ambiguïteit.
18. Een alternatieve vorm van de enquête zou erin kunnen bestaan dat de informanten een set van zelf geconstrueerde, goed gecontroleerde testzinnen moeten beoordelen. Op basis van de unanimiteit waarmee men de zinnen beoordeelt, kan dan de status van het criterium geëvalueerd worden. Een dergelijke benadering gaat er echter bij voorbaat van uit dat (i) dergelijke criteria bestaan en (ii) dat diegene die de enquête opstelt, weet welke criteria potentieel disambiguerend werken. Omdat o.i. nog nooit in een psycho-experimenteel design is nagegaan welke criteria disambiguerend werken, hebben we ervoor geopteerd om een stap terug te zetten en er niet zomaar van uit te gaan dat de criteria die in het verleden voorgesteld zijn, hoe plausibel ook, in het dagelijkse taalgebruik disambiguerend werken. In de plaats daarvan wilden we aan de hand van reëel taalmateriaal op inductieve wijze de criteria afleiden. Dat neemt niet weg dat het hier beschreven alternatief een waardevolle aanvulling zou kunnen bieden op de door ons gehanteerde methode, m.n. om op een iets grotere schaal de status van de gevonden criteria te verifiëren.
19. In drie gevallen kon 1 informant niet beslissen of de interpretatie adjectivisch of verbaal moest zijn (zin 9, 15 en 36 uit het eerste deel). In die gevallen was een uniformiteit van 50% uiteraard wel mogelijk.
20. Al moet daarbij onmiddellijk opgemerkt worden dat de modus (of andere subject- of taakgerelateerde factoren) waarin de enquête werd afgenomen hierin een niet onbelangrijke invloed zou kunnen hebben gespeeld: “[...] some cases where reading a sentence with normal intonation, as opposed to presenting it in written form, increased the number of acceptances” (Schütze 1996: 147).
21. Het ging om de volgende zinnen: zin 2, 4, 7, 10, 11, 13, 18, 19, 22, 23, 26, 28, 30, 34, 35, 41 van deel 1 en zin 2, 4, 6, 8, 9, 11, 16, 18 van deel 2.
22. De idee dat de frequentie van een item geïnterpreteerd kan worden in het kader van een betekenisontwikkeling is niet nieuw, ook in grammaticalisatiekringen wordt zoiets verondersteld (Hopper & Traugott 2003: 126-130).
23. Indien er twee identieke woordvormen gevonden werden, dan werd er gecontroleerd of die van woordsoort, lemma of frequentie verschillen:
 - Als de twee woordvormen op generlei wijze van elkaar verschillen, dan werden de frequenties van de woordvormen opgeteld. De frequentie van de types werd niet bij elkaar

opgeteld. Het is niet geheel duidelijk waarom Celex in die gevallen twee verschillende woordvormen heeft onderscheiden, gekoppeld aan hetzelfde type met dezelfde frequentie. In ieder geval is het zo dat de totale som van de frequenties van de woordvormen gelijk is aan de frequentie van het type.

- Als de woordsoort, het type en de frequentie van het type verschillen, maar de frequentie van de verschillende woordvormen is identiek, dan werden de frequenties van de woordvormen en de frequenties van de types opgeteld. Dat twee types onderscheiden worden in Celex heeft met de werkwoordelijke en adjectivische eigenschappen van het werkwoord in kwestie te maken. Alleen de frequenties van de woordvormen optellen en niet die van de types is ontoelaatbaar, omdat de totale frequentie van de woordvormen hoger zou zijn dan de hoogste frequentie van een van types.
 - Als de woordsoort, het type, de frequentie van het type en de frequentie van de woordvorm verschillen, dan werden de frequenties van de woordvormen en de frequenties van de types opgeteld. Alleen de frequenties van de woordvormen optellen en niet die van de types is ontoelaatbaar, omdat de totale frequentie van de woordvormen hoger zou zijn dan de hoogste frequentie van een van types. Dat is in het begin systematisch onderzocht en daarna voor elk vijfde item dat aan de criteria voldoet.
24. Door een dergelijke strenge procedure te hanteren, kunnen we min of meer zeker zijn dat de deelwoorden die een hoge graad van adjectiviteit vertonen, niet beïnvloed zijn door compilatiebiases en aldus representatief zijn voor het algemene taalgebruik.
 25. Dat betekent dat we gebruik maken van het regelsysteem dat de compilatoren opgesteld hebben om deelwoorden ofwel als werkwoord ofwel als adjectief te annoteren en dat we moeten vertrouwen op de accuratesse van de annotatieprocedure. Celex maakt gebruik van de *Woordenlijst van de Nederlandse Taal* en *Van Dale Groot Woordenboek van Hedendaags Nederlands* (Baayen et al. 1993). CGN (R5) (Van Eynde 2004) maakt gebruik van formele indicatoren zoals niet-werkwoordelijke stammen, adjectivische prefixen, graadsuffixen, buigings-s, naamvalssuffixen. Indien de deelwoorden niet aan een van die voorwaarden voldoen, dan wordt een werkwoordelijke interpretatie verondersteld, behalve wanneer een graadaanduidend bijwoord voorafgaat of als een deelwoord een andere valentie heeft dan de infinitief. Deelwoorden in de rode volgorde [AUX+PART] zijn altijd verbaal. Het spreekt voor zich dat de keuze van de compilatoren de adjectiviteits- en verbaliteitswaarde op een directe manier beïnvloeden, doordat er wellicht minder adjectieven gevonden zullen worden dan er misschien in realiteit zijn.
 26. Er werd in het CGN (R5) gezocht naar het deelwoord, gevolgd door een van de volgende tags: WW (vd) of ADJ (vrij, basis, zonder). Prenominaal gebruikte deelwoorden (*de betrokken persoon*), postnominale deelwoorden (*mensen geïnteresseerd in de opleiding*) en nominale deelwoorden (*de geïnteresseerde*) werden niet gebruikt.
 27. In de literatuur is men het niet eens over de bijwoorden die in aanmerking komen voor gradatie van het adjectief. Het meest opvallende punt van discussie is *erg*: sommigen menen dat *erg* gecombineerd moet worden met een adjectief, anderen vinden dat *erg* modificeerder kan zijn van

allerlei syntactische categorieën (zie bv. Klein 1998). Door dat gebrek aan eensgezindheid laten we *erg* terzijde, evenals *zo* waarvan de polysemie moeilijk controleerbaar is.

28. Een verbaal deelwoord wordt als een passief deelwoord beschouwd indien de vervoegde vorm van *zijn* vervangen kan worden door *worden* en de passieve zin teruggebracht kan worden naar de actieve pendant. Een verbaal deelwoord wordt als een voltooid deelwoord beschouwd indien de samengestelde tijd teruggebracht kan worden naar een enkelvoudige, onvoltooide tijd zonder dat de zin nonsensicaal of ongrammaticaal wordt.
29. Aangezien het enige resultaat in deze submodule '0' is, heeft het geen zin om een wegingsfactor toe te kennen.
30. De adjectiefpotentieelswaarden nemen reeds een cijfer tussen 0 en 1 aan, zodat die waarden niet meer getransformeerd dienden te worden. De adjectieftolerantiewaarden nemen daarentegen een cijfer tussen 0 en ∞ aan (het zijn nl. de absolute frequenties waarmee een deelwoord in een bepaalde morfosyntactische context voorkomt). De adjectiefpotentieelswaarden werden daarom getransformeerd, zodat ze eveneens variëren tussen 0 en 1: per tolerantie lijst werd het aantal voorkomens van een deelwoord gedeeld door het aantal voorkomens van het frequentste deelwoord in die lijst (het frequentste deelwoord in de lijst krijgt daardoor dus altijd waarde '1', de andere een waarde tussen 0 en 1). De EGVD-waarden, ten slotte, zijn binair (0 of 1), zodat een transformatie zich niet opdrong. Als bleek dat er voor een bepaalde bron geen cijfer voorhanden was, wegens te laag frequent of helemaal niet aanwezig, dan werd er in het geval van het adjectiefpotentieelsmeting de waarde '0.5' toegekend (precies tussen extreem adjectivisch en extreem werkwoordelijk in, dus onbeslist) en in het geval van de adjectieftolerantielijsten '0' (niet-aanwezigheid in de tolerantie lijsten is een argument pro verbaliteit).
31. Nu kan men opwerpen dat die resultaten te wijten zijn aan een gebrekkige operationalisatie van de status van het deelwoord. Het spreekt voor zich dat de resultaten in tabel 5.5 volledig afhangen van de manier waarop we de interpretatie van het deelwoord hebben trachten te operationaliseren en dat, met andere woorden, een andere invulling of een andere opbouw van het algoritme tot andere resultaten zou kunnen leiden. Het is dus zeker zo dat het algoritme in toekomstig onderzoek kritisch geëvalueerd en eventueel aangepast zal moeten worden, maar het zou ons wellicht te ver leiden om het algoritme als volledig ongeschikt te beoordelen. De volgende voorbeelden, die op basis van het algoritme als een historisch, adjectivisch, semi-formeel en adjectivisch gebruikt deelwoord geclassificeerd zijn, mogen verduidelijken dat de resultaten in tabel 5.5 niet zomaar af te schuiven zijn op een gebrekkige operationalisering, maar dat adjectivische deelwoorden in het hedendaagse taalgebruik in zekere mate in de rode volgorde [AUX+PART] kunnen voorkomen (de drie eerste voorbeelden worden nl. algemeen als adjectivisch beschouwd, over het vierde voorbeeld is discussie mogelijk):
 - (a) [...] dat opnieuw een meisje was vermist. (DS)
 - (b) [...] dat zijn boek op sensatie is belust. (DS)

(c) [...] dat zelfs het meest gestrenge bankgeheim niet is opgewassen tegen kwade trouw.
(DS)

(d) [...] dat de reis wel degelijk is afgelopen. (DS)

Dat zinnen als (a)-(d) in de rode volgorde [AUX+PART] kunnen verschijnen, hoeft al bij al niet te verbazen, aangezien het in een taalgemeenschap wel vaker voorkomt dat taalgebruikers zich om allerlei sociostilistische redenen van hypercorrecte vormen bedienen, maar dat het ook in een hoogstaande, kwalitatieve krant als *De Standaard* kan gebeuren, is tegen de achtergrond van de expliciete afkeuring in grammatica's en taaladviesboeken toch wel opmerkelijk. Uiteraard kan men beweren dat het daarbij om onoplettendheden gaat—de lage frequentie waarin adjectivische deelwoorden in de rode volgorde [AUX+PART] voorkomen lijkt voor die hypothese te spreken—maar dat neemt niet weg dat het taalsysteem blijkbaar in die omstandigheden in zekere mate variatie toelaat, zonder dat die opgemerkt wordt door of conflicteert met het taalgevoel of het taalnormenstelsel van een groep professionele taalgebruikers (de schrijver van het stuk en de eindredactie). Het is in dat verband wellicht niet toevallig dat naarmate de adjectivische deelwoorden duidelijker adjectivisch zijn (formeel > semi-formeel > adjectivisch gebruikt) de groene volgorde [PART+AUX] sterker staat, i.e. hoe minder twijfel er is over de adjectivische status van het deelwoord, hoe frequenter de groene volgorde [PART+AUX] gekozen wordt.

32. Daarmee willen we uiteraard niet gezegd hebben dat een derde optie onmogelijk is.
33. Het is niet ondenkbaar dat Pardoens hypothese—een deelwoord in de rode volgorde [AUX+PART] wordt dynamisch geïnterpreteerd—in feite niets met semantiek van doen heeft, maar met sociostilistische gemarkeerdheid. Het deelwoord in de rode volgorde [AUX+PART] staat nl. op een positie die afwijkt van wat gebruikelijk is in de dialecten en in het informele taalgebruik en het krijgt daardoor misschien zelfs wat extra beklemtoning, zodat het deelwoord extra in de kijker staat en zo de indruk kan ontstaan dat er een semantisch-aspectueel onderscheid is tussen de rode [AUX+PART] en de groene volgorde [PART+AUX].
34. Indien in een van de cellen de frequentie gelijk is aan 0, dan kan de LLR voor dat deelwoord niet berekend worden, omdat het natuurlijke logaritme van 0 gelijk is aan min oneindig en daardoor de LLR-score ook min oneindig is. Dat probleem hebben we trachten op te lossen door de nullen in de tabellen te vervangen door een erg klein getal: 1^{-24} . Op die manier zorgden we ervoor dat de LLR-berekening wel een interpreteerbare score opleverde (zonder dat de ingreep een beslissende invloed heeft op het feit of het deelwoord een statistisch significante collexeem is of niet).
35. Aangezien Pauwels nergens absolute cijfers geeft, kunnen we niet nagaan of de globale resultaten statistisch significant zijn.
36. Haeseryn (1990: 45) vraagt zich daarbij af of de analogiewerking niet in de omgekeerde richting kan werken, met name dat de rode volgorde [AUX+PART] in de twijfelgevallen verkozen wordt, omdat de rode volgorde [AUX+PART] in werkwoordelijke gezegdes in opmars is (vgl. ook ANS 1997: 1067-1069).

37. Even verderop presenteert De Schutter (1976: 238) de gegevens toch voor elk hulpwerkwoord apart (zonder er een nieuwe bespreking aan te wijden; de totalen daar verschillen licht van die op p. 232). Dezelfde tendensen als in de 1964-studie komen daarbij naar voren; het verschil tussen de vier categorieën (*hebben, zijn actief, zijn passief, worden*) is opnieuw significant ($\chi^2 = 23.48$, $df = 3$, $p < .0001$; eigen berekening).
38. Haeseryn zelf voert in dat verband geen statistische analyses uit, omdat hij het methodologisch onverantwoord vindt om resultaten uit verschillende types tests met elkaar te vergelijken.

Hoofdstuk 6

De psycholinguïstische dimensie

6.1 Inleiding

Nu de invloed van de contextuele, prosodische en semantische factoren op de keuze voor de groene [PART+AUX] of rode [AUX+PART] woordvolgorde vastgesteld is, ruimen we in het huidige hoofdstuk plaats voor wat wellicht de minst bekende dimensie is binnen het woordvolgordeonderzoek: de psycholinguïstische dimensie. Niettegenstaande dat het aantal psycholinguïstische onderzoeken naar de groene [PART+AUX] en rode [AUX+PART] woordvolgorde beperkt gebleven is tot een enkele studie, met name die van Hartsuiker & Westenberg (2000), heeft ze toch erg duidelijke en voor het huidige onderzoek belangwekkende resultaten opgeleverd. Hartsuiker & Westenberg gaan in hun studie nl. na of sprekers bij de keuze voor de rode [AUX+PART] of de groene [PART+AUX] woordvolgorde beïnvloed worden door wat in de psycholinguïstiek bekend staat als een *syntactisch persistentie*-effect. Zoals in paragraaf 3.2.1 al aangehaald werd, is syntactische persistentie de term die gebruikt wordt om te verwijzen naar de neiging van taalgebruikers om een bepaald syntactisch patroon dat recent geactiveerd is, hetzij actief, hetzij passief, te hergebruiken. Toegepast op de variatie in de tweeledige werkwoordelijke eindgroep betekent dat dan dat de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde bepaald wordt door de volgorde in een voorafgaande werkwoordelijke eindgroep (groen [PART+AUX] lokt groen [PART+AUX] uit, rood [AUX+PART] lokt rood [AUX+PART] uit). In paragraaf 6.2 willen we nagaan of er in onze dataset sporen gevonden kunnen worden van een dergelijk syntactisch persistentie-effect. Het moet echter van bij aanvang duidelijk zijn dat we aan de hand van ons corpusmateriaal

niet in staat zullen zijn om directe evidentie te bieden voor een dergelijk effect, en dat de interpretatie in psycholinguïstische termen dan ook slechts voorwaardelijk is en in de toekomst verder geschraagd zal moet worden door psycho-experimenteel onderzoek.

Diezelfde opmerking geldt ook voor de tweede (nieuw geïntroduceerde) factor die we in dit hoofdstuk zullen bestuderen, met name de frequentie van het deelwoord. Tegen de achtergrond van de theoretische doelstelling van dit proefschrift, de syntactische variatie in de werkwoordelijke eindgroep inkaderen in een emergente¹ visie op taal, is de selectie van de factor *frequentie* evident. De frequentie waarmee een taaleenheid gebruikt wordt, bepaalt nl. onder meer de manier waarop een taaleenheid afgebakend wordt in het geheugen, de mate waarin een taaleenheid toegankelijk is en de snelheid waarmee een taaleenheid verwerkt wordt in taalproductie- en taalperceptieprocessen, waardoor het niet ondenkbaar is dat de frequentie van een taaleenheid, i.c. het deelwoord, ook een rol zou kunnen spelen bij de keuze voor de rode [AUX+PART] of groene [PART+AUX] volgorde. Of en in welke mate de frequentie van het deelwoord een invloed uitoefent op de keuze tussen een van de woordvolgordevarianten komt in paragraaf 6.3 ter sprake. In paragraaf 6.4 wordt tot slot nog een samenvatting gegeven van de verwezenlijkingen in dit hoofdstuk.

6.2 De impact van syntactische persistentie

In deze paragraaf bespreken we het effect van syntactische persistentie op de keuze voor de groene [PART+AUX] of rode [AUX+PART] woordvolgorde. Eerst geven we een overzicht van de belangrijkste literatuur ter zake (paragraaf 6.2.1), vervolgens worden de verschillende hypotheses voorgesteld (paragraaf 6.2.2) en tot slot presenteren en evalueren we de resultaten van ons onderzoek (paragraaf 6.2.3). We proberen daarbij een antwoord te geven op de vraag of de volgorde van een voorafgaande tweeledige (paragraaf 6.2.3.1) of meerledige (paragraaf 6.2.3.2) werkwoordelijke eindgroep een effect heeft op de volgorde in de targeteindgroep. Daarnaast kijken we ook of de afstand tussen de targeteindgroep en de voorafgaande eindgroep een invloed uitoefent op de mate waarin een voorafgaande eindgroep een

effect heeft (6.2.3.3) en komen we terug op het in hoofdstuk 3 besproken syntactisch persistentie-effect in het onderzoek van Haeseryn (1990) (paragraaf 6.2.3.4).

6.2.1 *Status quaestionis*

Zoals in de inleiding al aangekondigd werd, is de invloed van syntactische persistentie op de keuze voor de groene [PART+AUX] of rode [AUX+PART] volgorde totnogtoe enkel bestudeerd door Hartsuiker & Westenberg (2000).² Voor de duidelijkheid herhalen we dat syntactische persistentie de neiging aanduidt van taalgebruikers om een bepaald syntactisch patroon (zoals de groene [PART+AUX] of rode [AUX+PART] volgorde) dat recent geactiveerd is, hetzij actief, hetzij passief, te hergebruiken. Hartsuiker & Westenberg onderzochten de invloed van syntactische persistentie aan de hand van een experimenteel design waarbij 66 testsubjecten een matrixzin en een bijzin zonder eindgroep aangeboden kregen. De testsubjecten werd daarbij gevraagd om de zinnen verder aan te vullen. Een deel van die zinnen, de primezinnen, werd zo opgesteld dat de informanten slechts één woordvolgorde konden kiezen, i.e. de rode [AUX+PART] of de groene [PART+AUX]: naast de matrixzin en het niet-verbale gedeelte van de bijzin werd ook het hulpwerkwoord of het deelwoord aangeboden, zodat de testsubjecten slechts konden aanvullen met respectievelijk een deelwoord (het (a)-voorbeeld; de rode volgorde [AUX+PART]) of een hulpwerkwoord (het (b)-voorbeeld; de groene volgorde [PART+AUX]) (de voorbeelden komen uit Hartsuiker & Westenberg 2000: B34):

(45a) Ik kon er niet door omdat de weg was...

(45b) Ik kon er niet door omdat de weg geblokkeerd...

Een ander deel van de zinnen, de doelzinnen, werd niet aan een dergelijke beperking onderworpen, zodat de testsubjecten daar de vrije keuze hadden tussen een groene [PART+AUX] of een rode [AUX+PART] volgorde:

(46) Jan vertelde de inspecteur dat hij niets...

Door de doelzinnen afwisselend vooraf te laten gaan door een primezin met de groene [PART+AUX] volgorde, door een primezin met de rode volgorde [AUX+PART] en door een primezin die geen werkwoordelijke eindgroep, maar een NP uitlokt, trachtten

Hartsuiker & Westenberg te achterhalen of testsubjecten zich bij de selectie voor een van de woordvolgordes in de doelzinnen laten beïnvloeden door de woordvolgorde in de primezinnen. Elke informant kreeg (24 prime-doelparen =) 48 experimentele items (de helft met *zijn* en de helft met *hebben*) voorgeschoteld, verspreid over 192 fillers, waardoor de subjecten zich niet bewust waren van het opzet van de studie. Het onderzoek werd zowel in geschreven als gesproken modus uitgevoerd. De resultaten van het onderzoek maken het volgende duidelijk:

- Informanten gebruiken significant vaker de groene volgorde [PART+AUX] in de doelzinnen als de groene volgorde [PART+AUX] in de primezinnen uitgelokt werd en de rode volgorde [AUX+PART] als de rode volgorde [AUX+PART] uitgelokt werd. Dat betekent dat de keuze voor een van de woordvolgordes bepaald wordt door de woordvolgorde in een onmiddellijk voorafgaande zin.
- Er is geen verschil tussen gesproken en geschreven taalgebruik, wat erop wijst dat de syntactische persistentie in de beide modi werkzaam is. Dat is des te opvallender, daar een productieproces in de geschreven modus trager is en er daardoor meer tijd voorhanden is om eventuele woordvolgordes in de loop van het productieproces te veranderen (zie echter ook Pickering et al. 2000: 206 voor andere referenties).

In het vervolg van deze paragraaf zullen we aan de hand van ons corpusmateriaal (zie paragraaf 4.1, tabel 4.2) de syntactische-persistentie-hypothese verifiëren. We zullen meer bepaald nagaan of een werkwoordelijke eindgroep die zich in de voorgaande, artikelinterne context bevindt van de zinnen die we als relevant beschouwd hebben (verder: de targetzinnen), de keuze voor een van de syntactische alternatieven in de targeteindgroep bepaald heeft. Dat Hartsuiker & Westenberg ontdekt hebben dat een persistentie-effect niet alleen in gesproken taal, maar ook in geschreven taal voorkomt, doet het beste verhoppen voor de resultaten van ons onderzoek, maar toch moet hier gewezen worden op een aantal verschilpunten met het onderzoek van Hartsuiker & Westenberg die van belang zouden kunnen zijn voor de graad van succes van onze analyses. Zo waren de testsubjecten in het onderzoek van Hartsuiker & Westenberg studenten (i) waarvan verondersteld kan worden dat ze niet professioneel met taal bezig zijn; (ii) wier taaluitingen redactioneel niet gecontroleerd werden; (iii) die onder

een zekere mate van tijdsdruk stonden; (iv) die aangeraden werd zo spontaan mogelijk te schrijven. In onze dataset van zinnen uit krantenartikelen is er daarentegen sprake van professionele taalgebruikers, van een ruime mate van redactionele controle en een beperkte tijdsdruk, waardoor het schrijfproces bewuster, gecontroleerder en trager verloopt (zie de resultaten in hoofdstuk 3). Dat kan ertoe leiden dat onderliggende psycholinguïstische processen van structurele priming bij journalisten doorkruist worden doordat bijvoorbeeld om stilistische redenen afgewisseld wordt tussen de groene [PART+AUX] en de rode [AUX+PART] volgorde (zie bv. de richtlijnen van Heidbuchel 1967; paragraaf 1.3.1.2). Indien, met andere woorden, zou blijken dat de corpusgebaseerde resultaten van ons onderzoek niet dezelfde tendensen vertonen als de psycholinguïstische van Hartsuiker & Westenberg, dan kan dat vermoedelijk toegeschreven worden aan de omstandigheden van het productieproces dat voorafgegaan is aan de beide types taaluitingen. Maar zelfs indien onze resultaten zouden overeenkomen met die van Hartsuiker & Westenberg, dan nog bestaat de mogelijkheid dat er niet zozeer een psycholinguïstisch persistentie-effect werkzaam is, maar wel een stilistisch persistentie-effect (deze of gene redacteur heeft een persoonlijke voorkeur voor de groene [PART+AUX] of de rode [AUX+PART] volgorde en gebruikt die volgorde bewust in zijn of haar artikelen).

6.2.2 Hypotheses

Op basis van het onderzoek van Hartsuiker & Westenberg (2000) formuleren we de volgende hypothese:

- [HYP 6.1] De keuze voor de groene [PART+AUX]/rode [AUX+PART] volgorde stijgt, wanneer een werkwoordelijke eindgroep in de groene [PART+AUX]/rode [AUX+PART] volgorde voorafgaat (paragraaf 6.2.3.1 en 6.2.3.2).

Gegeven het psycholinguïstische karakter van syntactische persistentie en gegeven het feit dat de mogelijkheden van het menselijke geheugen niet onbeperkt zijn (zie bv. Gibson 1998; zie noot 2), verwachten we dat het persistentie-effect vermindert naarmate meer tijd verstrijkt:

[HYP 6.2] Op voorwaarde dat [HYP 6.1] bevestigd kan worden, daalt de keuze voor de groene [PART+AUX]/rode [AUX+PART] volgorde, naarmate de afstand tot de voorafgaande groene [PART+AUX]/rode [AUX+PART] werkwoordelijke eindgroep groter wordt (paragraaf 6.2.3.3).

6.2.3 Resultaten en discussie

Voor het onderzoek naar de neiging van een werkwoordelijke eindgroep om dezelfde vorm aan te nemen als die van een voorafgaande werkwoordelijke eindgroep, zullen we de voorafgaande context van de targeteindgroepen in onze dataset afspeuren naar voorafgaande werkwoordelijke eindgroepen (verder: pre-eindgroepen). Daarbij zijn de volgende principes gehanteerd:

1. De pre-eindgroepen die in het vervolg van deze paragraaf bestudeerd worden, mogen maximaal 10 artikel-interne zinnen van de targeteindgroep verwijderd zijn. Indien er meer dan een eindgroep in de voorafgaande context voorkwam, werd enkel die eindgroep geselecteerd die zich het dichtst bij de targeteindgroep bevond.
2. De selectie van de pre-eindgroepen was, in tegenstelling tot wat het geval was bij de targeteindgroepen (zie hoofdstuk 2), niet aan andere voorwaarden gebonden dan de net genoemde. De reden daarvoor is dat het opleggen van selectiecriteria aan de pre-eindgroepen tot niets dient: laten we even in herinnering brengen dat de selectiecriteria waaraan de targeteindgroepen moesten voldoen, opgelegd werden om een aantal factoren die potentieel een rol zouden kunnen spelen, te controleren, zodat we er bij de interpretatie van de resultaten gerust op konden zijn dat die factoren de resultaten niet vertekenen. Voor de studie van de pre-eindgroep heeft een dergelijke controle geen zin, omdat we niet te weten willen komen welke factor de woordvolgorde bepaald heeft, we willen enkel het resultaat van dat keuzeproces gebruiken als verklarende factor voor de woordvolgordeselectie van de daaropvolgende eindgroep.³

3. De pre-eindgroepen die in paragraaf 6.2.3.1 bestudeerd worden, moeten tweeledig zijn en bestaan uit een deelwoord en een vervoegde vorm van *zijn*, *hebben* of *worden*.
4. De pre-eindgroepen die in paragraaf 6.2.3.2 bestudeerd worden, moeten minstens tweeledig zijn, een deelwoord en het hulpwerkwoord *zijn*, *hebben* of *worden* bevatten en een reële en systematische variatiemogelijkheid inhouden.⁴ Die uitbreiding houdt in dat ook pre-eindgroepen in hoofdzinnen geanalyseerd zijn, althans voor zover ze minstens driedelig zijn (in dat geval staan nl. twee infinitieve werkwoorden in de tweede zinspool, die onderling van plaats kunnen wisselen; het finiete werkwoord in de eerste zinspool is in een dergelijke context niet verplaatsbaar).

6.2.3.1 Voorafgaande tweeledige werkwoordelijke eindgroepen

In een eerste stap verifiëren we de invloed die de woordvolgorde in een tweeledige pre-eindgroep, die bestaat uit het hulpwerkwoord *zijn*, *hebben* of *worden* en een deelwoord en zich in een bijzin bevindt, heeft op de volgorde in de targeteindgroep. We laten daarbij de afstand tussen de beide eindgroepen voorlopig nog buiten beschouwing. De hypothese is de volgende:

[HYP 6.1] De keuze voor de groene [PART+AUX]/rode [AUX+PART] volgorde stijgt, wanneer een werkwoordelijke eindgroep in de groene [PART+AUX]/rode [AUX+PART] volgorde voorafgaat.

Tabel 6.1: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type tweeledige pre-eindgroep (in een bijzin).

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Pre-eindgroep: groene volgorde [PART+AUX]	45.76 % (151/330)	54.24 % (179/330)
Pre-eindgroep: rode volgorde [AUX+PART]	24.08 % (189/785)	75.92 % (596/785)

Van de 2390 eindgroepen in onze dataset worden er 1115 voorafgegaan door een tweeledige werkwoordelijke cluster, dat is 46.64% van het totaal (1115/2390). 29.59% van de observaties die voorafgegaan worden door een tweeledige eindgroep is een groene volgorde [PART+AUX] (330/1115), 70.40% is een rode volgorde [AUX+PART] (785/1115). Als men in het achterhoofd houdt dat de globale proporties van de groene [PART+AUX] en rode [AUX+PART] volgorde in de targeteindgroepen die voorafgegaan worden door een tweeledige werkwoordelijke eindgroep respectievelijk 30.49% en 69.51% bedraagt, dan laten de frequentieverhoudingen in tabel 6.1 duidelijk zien dat resultaten van Hartsuiker & Westenberg (2000) aan de hand van onze data bevestigd kunnen worden: de voorkeur voor de rode volgorde [AUX+PART] in de targeteindgroep stijgt, indien de tweeledige pre-eindgroep in de rode volgorde [AUX+PART] voorkomt (75.92%; verschil met de globale proportie rode volgordes [AUX+PART] in tabel 6.1: 6.41%), terwijl een groene volgorde [PART+AUX] in de targeteindgroep verkozen wordt indien een groene volgorde [PART+AUX] voorafgaat (45.76%; verschil met de globale proportie groene volgordes [PART+AUX] in tabel 6.1: 15.27%). Het verschil in voorkeur is statistisch significant ($\chi^2 = 51.53$, $df = 1$, $p < .0001$). De o.r. van 2.66 (B.I. = 2.01-3.52) geeft aan dat de odds voor de groene volgorde [PART+AUX] gegeven een groene pre-eindgroep [PART+AUX] meer dan 2.5 keer groter zijn dan de odds voor een groene volgorde [PART+AUX] gegeven een rode pre-eindgroep (in de volgende paragraaf betrekken we de eindgroepen zonder pre-eindgroep in de analyse).

Daaruit blijkt, met andere woorden, ondanks het feit dat het type geschreven taal verschilt van het type dat in Hartsuiker & Westenberg (2000) onderzocht is, dat er inderdaad een syntactisch persistentie-effect lijkt te zijn dat ervoor zorgt dat de keuze voor een van de volgordes geassocieerd is met de volgorde in een tweeledige pre-eindgroep (bevestiging [HYP 6.1]; zie echter de opmerking die we aan het eind van paragraaf 6.2.1 maakten).

6.2.3.2 *Voorafgaande twee- en meerledige werkwoordelijke eindgroepen*

De vraag die we ons vervolgens zullen stellen, is in hoeverre het net vastgestelde persistentie-effect ook geldt voor tweeledige eindgroepen die voorafgegaan worden door andere types werkwoordelijke eindgroepen, meer bepaald twee- of meerledige eindgroepen in hoofdzinnen of bijzinnen met minstens een deelwoord en het hulpwerkwoord *hebben*, *zijn* of *worden*. Puur infinitivische clusters (zoals [INF/AUX]) zijn niet in de analyse betrokken, omdat het type variatie er toch van een andere orde is (de rode volgorde [AUX+INF] is er extreem dominant) en we er derhalve momenteel nog niet zomaar vanuit kunnen gaan dat de volgorde in zuiver infinitivische clusters als prime kan gelden voor de volgorde in eindgroepen met een deelwoord; daarvoor is bijkomend onderzoek nodig.

In tabel 6.2 worden de verschillende types pre-eindgroepen die we in onze dataset gevonden hebben verdeeld over de responsvariabele. Het gaat om de volgende types: [AUX+PART] (voorbeeldzin (47)), [teINF+PART] (voorbeeldzin (48)), [AUX/INF+PART] (voorbeeldzin (49)), [AUX+INF+PART] (voorbeeldzin (50)), [AUX+PART+INF] (voorbeeldzin (51)), [PART+AUX] (voorbeeldzin (52)), [PART+teINF] (voorbeeldzin (53)), [AUX/PART+INF] (voorbeeldzin (54)) en [PART+AUX+INF] (voorbeeldzin (55)). Een '+'-teken in het formele notatiesysteem betekent overigens dat de werkwoordelijke elementen in elkaars onmiddellijke buurt staan, de slash (/) betekent daarentegen dat er tussen de verbale elementen tussenliggend materiaal kan komen.

- (47) Een vliegtuig van American Airlines heeft een noodlanding gemaakt nadat een van de twee motoren was uitgevallen. (DS)
- (48) Daarin worden Haarlemse inspecteurs ervan verdacht bewust drugstransporten door de vingers te hebben gezien. (DS)

- (49) De opslag van de embryo's kon worden verlengd als zowel de donor van de zaadcellen als die van de eicel dat wilden. (DS)
- (50) Dat is weer zo'n misbruik dat hoogdringend moet worden afgeschaft! (DS)
- (51) Hebron, dat al in maart grotendeels moest ontruimd zijn, blijft bezet. (DS)
- (52) Zijn er dan de voorbije weken of maanden moeilijkheden geweest die tot het ontslag van informatici geleid hebben? (DS)
- (53) Onder een schuilnaam boekte Wagner twee kaartjes voor Siegfried bij een reisbureau, dat toegang de kleinoden 'op de tweedehandsmarkt' op de kop getikt te hebben. (DS)
- (54) In samenwerking met het cultureel centrum kan een bescheiden programmatie opgezet worden. (DS)
- (55) De Palestijnse onderhandelaar, Saeb Erekat, zei dat er nog een aantal obstakels over waren, maar dat er mogelijk later vooruitgang gemeld kan worden.

Alvorens tot een bespreking van de tabel en de bijbehorende analyses over te gaan, moet eerst nog opgemerkt worden dat een aantal van de pre-eindgroepen zo infrequent is, dat ze niet in de tabel opgenomen is en niet betrokken is in de statistische analyses (wegens onbetrouwbaar). Het ging meer bepaald om [PART+AUX+PART] (n = 6), [AUX/INF+INF+PART] (n = 5), [AUX+INF+INF+PART] (n = 3), [AUX+teINF+INF+PART] (n = 1), [AUX/INF+PART+INF] (n = 1), [AUX/PART+INF+INF] (n = 6), [AUX+INF+PART+INF] (n = 1), [AUX+PART+INF+INF] (n = 1) en [AUX+PART+INF+INF] (n = 1).

Tabel 6.2: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type twee- of meerledige pre-eindgroep.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Pre-eindgroep: [AUX+PART]	24.08 % (189/785)	75.92 % (596/785)
Pre-eindgroep: [teINF+PART]	21.05% (8/38)	78.95% (30/38)
Pre-eindgroep: [AUX/INF+PART]	19.63% (21/107)	80.37% (86/107)
Pre-eindgroep: [AUX+INF+PART]	24.68% (19/77)	75.32% (58/77)
Pre-eindgroep: [AUX+PART+INF]	20% (4/20)	80% (16/20)
Pre-eindgroep: [PART+AUX]	45.76 % (151/330)	54.24 % (179/330)
Pre-eindgroep: [PART+teINF]	35.71% (5/14)	64.29% (9/14)
Pre-eindgroep: [AUX/PART+INF]	57.33% (43/75)	42.67% (32/75)
Pre-eindgroep: [PART+AUX+INF]	50% (14/28)	50% (14/28)

De uitbreiding naar andere types pre-eindgroepen maakt duidelijk dat van de 2390 eindgroepen in onze dataset er 1499 voorafgegaan worden door een of ander type werkwoordelijke eindgroep (met minstens een deelwoord en een vervoegde vorm van *zijn*, *hebben* of *worden*), i.e. 62.72% van het totaal. De uitbreiding maakt verder ook duidelijk dat er 384 pre-eindgroepen in de dataset zitten die niet tweeledig zijn en bestaan uit een deelwoord en het hulpwerkwoord *zijn*, *hebben* of *worden*. Belangrijker is echter dat 30.80% van de observaties die voorafgegaan worden door een of ander type eindgroep in de groene volgorde [PART+AUX] voorkomt (454/1474⁵), 69.20% in de rode volgorde [AUX+PART] (1020/1474). Als we dat vergelijken met de globale volgordevoorkeur bij de gevallen met een tweeledige pre-eindgroep die bestaan uit een deelwoord en het hulpwerkwoord *zijn*, *hebben* of *worden* (m.n. 30.49%

groen [PART+AUX] en 60.51% rood [AUX+PART]), dan zijn de verschillen niet significant ($\chi^2 = 0.44$, $df = 1$, $p > .05$).

Een chi-kwadraatanalyse maakt ook voor de gegevens in tabel 6.2 duidelijk dat er een significant effect is van type pre-eindgroep op de keuze van de woordvolgorde in de targeteindgroep ($\chi^2 = 91.48$, $df = 8$, $p < .0001$).⁶ Het is echter niet onmiddellijk duidelijk hoe we dat effect precies moeten begrijpen. Om de aard van het effect beter te begrijpen, voeren we daarom een residuanalyse en een gepartitioneerde chi-kwadraatanalyse uit. Met behulp van een residuanalyse gaan we eerst na of er een bepaald type pre-eindgroep aangewezen kan worden dat meer dan andere types heeft bijgedragen tot de globale sterkte van het effect. In ons geval blijken dat voornamelijk het [AUX+PART]-patroon (residu = 5.97), het [PART+AUX]-patroon (residu = 6.68) en het [AUX/INF+PART]-patroon (residu = 5.11) te zijn: het eerste patroon stimuleert vaker dan verwacht de rode volgorde [AUX+PART] in de targeteindgroep, de andere patronen stimuleren de groene volgorde [PART+AUX] in de targeteindgroep. De andere types wijken nauwelijks af van wat verwacht wordt (residu < 3). Uit de gepartitioneerde chi-kwadraatanalyse, die erop gericht is om het aantal categorieën in een tabel met meer dan 2 categorieën te reduceren door groepen samen te nemen die statistisch niet-significant van elkaar verschillen, blijkt dat de eerste vier types en de laatste vier types niet statistisch significant verschillen in hun effect op de woordvolgorde in de targeteindgroep en daarom samengenomen kunnen worden. Het vijfde type is een twijfelgeval, omdat het zowel bij de eerste als bij de laatste groep gevoegd kan worden, zonder dat het er significant van verschilt. We zullen ervoor opteren om het vijfde type bij de tweede groep te voegen, omdat op die manier twee inhoudelijk homogene groepen ontstaan: de eerste groep bevat die types waarvan het deelwoord rechts staat van het eerste bepalende werkwoord (i.e. het hulpwerkwoord dat het minst snel de functie van persoonsvorm overneemt als je een meerledige cluster terugbrengt naar een minder gelede cluster; Van der Lubbe 1965: 274-275), terwijl de tweede groep de types bevat waarvan het deelwoord links staat van het eerste bepalende hulpwerkwoord. Op basis van dat verschil in positie van het deelwoord ten opzichte van het eerste bepalende werkwoord zouden we bij uitbreiding kunnen spreken van de rode [AUX+PART] (deelwoord rechts) en de groene [PART+AUX] (deelwoord links) woordvolgorde:

Tabel 6.3: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type pre-eindgroep (twee- en meerledige eindgroepen, samengevat in twee basistypes).

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Pre-eindgroep: groene volgorde [PART+AUX]	46.75 % (223/477)	53.25 % (254/477)
Pre-eindgroep: rode volgorde [AUX+PART]	23.68 % (242/1022)	76.32 % (780/1022)

Tabel 6.3 laat ongeveer hetzelfde zien als tabel 6.1, met name dat een voorafgaande rode werkwoordelijke cluster [AUX+PART], vergeleken met het globale aantal rode volgordes [AUX+PART] (69.20%), een belangrijke toename van rode volgordes [AUX+PART] in de targeteindgroep teweegbrengt (76.32%). Omgekeerd geldt dat een voorafgaande groene woordvolgorde [PART+AUX] een aanzienlijke stijging van de groene volgorde [PART+AUX] laat zien in de targetcluster (46.75%), vergeleken met het globale aantal groene volgordes [PART+AUX] (30.80%). Het verschil tussen die beide groepen is erg significant ($\chi^2 = 80.9$, $df = 1$, $p < .0001$) en de o.r. van 2.83 (B.I. = 2.23-3.59) geeft aan dat de odds voor de groene volgorde [PART+AUX] gegeven een groene pre-eindgroep [PART+AUX] meer dan 2.5 keer groter zijn dan de odds voor een groene volgorde [PART+AUX] gegeven een rode pre-eindgroep [AUX+PART].

Ook wanneer andere types pre-eindgroepen in de analyse betrokken worden, blijkt er met andere woorden een syntactisch persistentie-effect te zijn, waardoor de keuze voor de volgorde afhangt van een eerder gemaakte keuze (bevestiging van [HYP 6.1]; zie echter de opmerking die we aan het eind van paragraaf 6.2.1 maakten).⁷

Maar wat gebeurt er wanneer de taalgebruikers in de voorafgaande context geen dergelijke keuze hebben moeten maken?⁸ In dat geval blijken de proporties groene [PART+AUX] en rode [AUX+PART] woordvolgorde zich ongeveer tussen de geïntensifieerde voorkeur voor rood [AUX+PART] (voor de gevallen met een rode pre-eindgroep [AUX+PART]) en de geïntensifieerde voorkeur voor groen

[PART+AUX] (voor de gevallen met een groene pre-eindgroep [PART+AUX]) te bevinden: 36.36% voor de groene volgorde [PART+AUX] (n = 324) en 63.64% voor de rode volgorde [AUX+PART] (n = 567). Het verschil tussen die drie categorieën is erg significant ($\chi^2 = 85.49$, $df = 2$, $p < .0001$) en geen van de drie categorieën kan samengenomen worden met een andere categorie (na een gepartitioneerde chikwadraatanalyse).⁹ Dat betekent, met andere woorden, dat taalgebruikers de keuze voor een van de twee woordvolgordealternatieven in de targeteindgroep laten afhangen van de woordvolgorde in een voorafgaande eindgroep—psycholinguïstisch geformuleerd: de neiging hebben om de woordvolgorde die recent geactiveerd is te hergebruiken. Indien er recent geen activatie was van het syntactische patroon, zakt de voorkeur voor een van de volgordes terug tot op het niveau van wat gemiddeld genomen gebruikelijk is (over de hele dataset gezien komt de groene volgorde [PART+AUX] 33.01% keer voor (789/2390), in eindgroepen zonder pre-eindgroep is dat, zoals gezegd, 36.36% (324/891); het verschil is niet-significant: $\chi^2 = 3.25$, $df = 1$, $p > 0.05$).

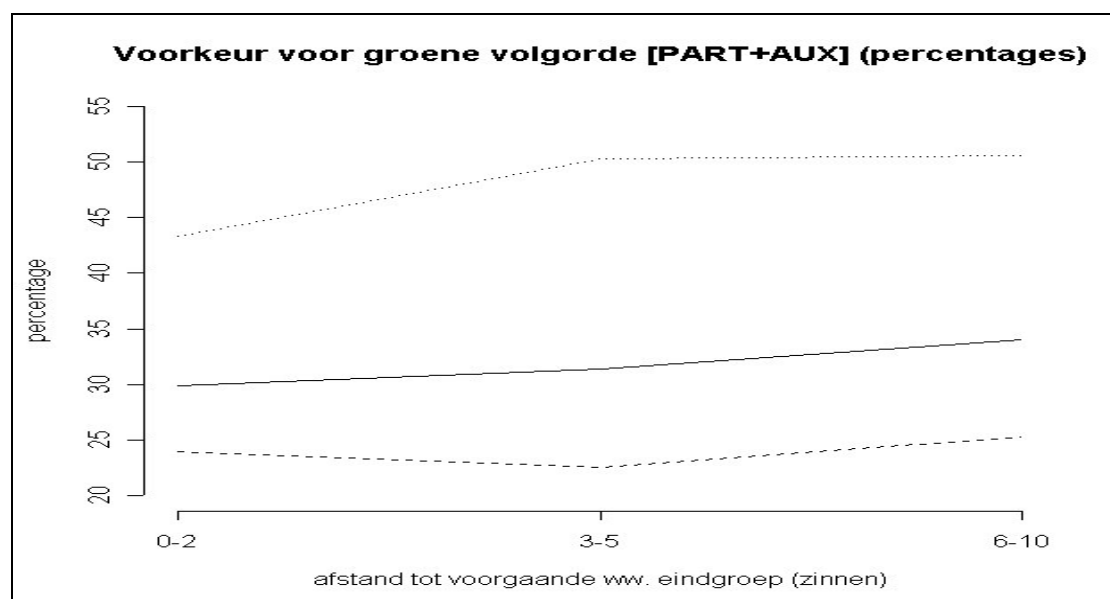
6.2.3.3 *De activatiegraad van voorafgaande werkwoordelijke eindgroepen*

De vraag die men zich nu nog kan stellen, is in hoeverre de afstand tussen de beide eindgroepen de kracht van het primingeffect beïnvloedt. Als we voorlopig aannemen dat het effect van een pre-eindgroep op een psycholinguïstische manier geïnterpreteerd mag worden, i.e. taalgebruikers hebben de (onbewuste) neiging om de volgorde van een eindgroep te recyclen, omdat dat de computationele last die gepaard gaat met het taalproductieproces reduceert (ze kunnen nl. dezelfde volgorde gebruiken als die die recent geactiveerd werd; Pickering et al. 2000: 210; zie ook noot 2), dan is het niet onaannemelijk, gezien de beperkingen op het menselijke geheugen (bv. Gibson 1998), dat de activatie van een bepaalde volgorde niet permanent is, maar afneemt naarmate meer tijd verstrijkt. De verwachting is dan ook de volgende:

[HYP 6.2] De keuze voor de groene [PART+AUX]/rode [AUX+PART] volgorde daalt, naarmate de afstand tot de voorafgaande groene [PART+AUX]/rode [AUX+PART] werkwoordelijke eindgroep groter wordt.

Omdat het onmogelijk is aan de hand van corpusmateriaal na te gaan hoe het effect van structurele persistentie beïnvloed wordt door de tijd, zullen we *tijd* hier operationaliseren als het aantal zinnen dat zich tussen de beide eindgroepen bevindt: meer zinnen staat voor meer tijd (een zin is opgevat als het groepje woorden dat zich tussen twee leestekens bevindt die het einde van een zin aanduiden, i.e. een punt, een uitroepeteken of een vraagteken). In figuur 6.1 wordt het effect van een voorafgaande werkwoordelijke eindgroep (twee- en meerledige eindgroepen, samengevat in twee basistypes; zie tabel 6.3) uitgesplitst over de afstand tussen de twee werkwoordelijke clusters, met name 0 tot 2 zinnen, 3 tot 5 zinnen en 6 tot 10 zinnen. De volle lijn in figuur 6.1 representeert de voorkeur voor de groene volgorde [PART+AUX] onafhankelijk van het type pre-eindgroep, de streeplijn geeft de voorkeur voor de groene volgorde [PART+AUX] weer voor de gevallen die voorafgegaan worden door een werkwoordelijke eindgroep in de rode volgorde [AUX+PART] en de stippellijn staat voor de voorkeur voor de groene volgorde [PART+AUX] voor de gevallen die voorafgegaan worden door een werkwoordelijke eindgroep in de groene volgorde [PART+AUX] (de voorkeur voor de rode volgorde [AUX+PART] is telkens het spiegelbeeld).

Figuur 6.1: Voorkeur voor de groene volgorde [PART+AUX], uitgesplitst per lengtecategorie, onafhankelijk van het type pre-eindgroep (volle lijn), voor de groene pre-eindgroepen [PART+AUX] (stippellijn) en de rode pre-eindgroepen [AUX+PART] (streeplijn).



Uit figuur 6.1 blijkt dat er geen noemenswaardige stijging of daling in voorkeur is, hetgeen bevestigd wordt door een Mantel-Haenszel chi-kwadraattoets (onafhankelijk van type pre-eindgroep: $MH-\chi^2 = 1.55$, $df = 1$, $p > .05$; rode pre-eindgroep [AUX+PART]: $MH-\chi^2 = 0.49$, $df = 1$, $p > .05$; groene pre-eindgroep [PART+AUX]: $MH-\chi^2 = 1.99$, $df = 1$, $p > .05$). In tegenstelling tot de verwachting worden er dus niet meer groene woordvolgordes [PART+AUX] gebruikt naarmate de afstand tot de voorgaande rode werkwoordelijke cluster [AUX+PART] groter wordt en, omgekeerd, worden er niet meer rode woordvolgordes [AUX+PART] gebruikt naarmate de afstand tot de voorgaande groene werkwoordelijke cluster [PART+AUX] stijgt (falsificatie van [HYP 6.2]). Een gestratificeerde analyse met woordvolgorde in de targeteindgroep als responsvariabele, woordvolgorde in de pre-eindgroep als verklarende variabele en de afstand tussen de twee eindgroepen als controlevariabele laat, zoals verwacht, zien dat er een effect is van pre-eindgroep, onafhankelijk van de afstand tot de targeteindgroep ($CMH-\chi^2 = 80.23$, $df = 1$, $p < .0001$; globale o.r. = 2.82, B.I. = 2.24-3.55; $BR-D-\chi^2 = 1.87$, $df = 2$, $p > .05$; geen interactie- of redundantie-effecten in de partiële tabellen).

De vraag die daaruit voortvloeit, is dan uiteraard in hoeverre het houdbaar is om het effect psycholinguïstisch te interpreteren. Een alternatieve interpretatie (zie ook de opmerking die we aan het eind van paragraaf 6.2.1 maakten) die niet uit te sluiten valt, is dat deze of gene redacteur een persoonlijke voorkeur heeft voor de groene [PART+AUX] of de rode volgorde [AUX+PART] en die volgorde dan ook bewust gebruikt in zijn of haar artikelen. Van een psycholinguïstisch persistentie-effect is dan uiteraard geen sprake meer, wel van een stilistisch persistentie-effect. Welke van de beide interpretaties correct is, is niet eenvoudig uit te maken. In ieder geval kan wel nog opgemerkt worden dat het uitblijven van een afname van het persistentie-effect naarmate meer tijd verstrijkt, niet automatisch hoeft te betekenen dat een psycholinguïstische interpretatie uitgesloten is, aangezien sommige resultaten van goed gecontroleerde psycholinguïstische experimenten laten zien dat structurele priming over verscheidene zinnen heen actief kan zijn; andere resultaten maken dan weer duidelijk dat priming enkel werkzaam is als de prime- en de targetzin in elkaars onmiddellijke buurt staan (Pickering et al. 2000: 212-213). Hoe het ook moge zijn, het is duidelijk dat de interpretatie van de resultaten in deze en de vorige twee paragrafen

vooral nog in voorzichtige termen gesteld moet worden. Toekomstig onderzoek zal hierover klaarheid moeten scheppen.

6.2.3.4 *Onbedoelde syntactische priming in psycho-experimenteel onderzoek*

Tot slot van deze paragraaf over syntactische persistentie willen we nog even terugkomen op de hypothese die we in paragraaf 3.2.1 met betrekking tot het onderzoek van Haeseryn (1990) formuleerden, met name dat een deel van de resultaten daar gecorrumpeerd zouden kunnen zijn door een syntactisch persistentie-effect. We herhalen kort dat Haeseryn aan de hand van een aantal gedragsexperimenten het effect van een set van (socio-) linguïstische factoren op de keuze voor de groene [PART+AUX] of de rode woordvolgorde [AUX+PART] wou testen. Een van de tests die hij daarvoor gebruikt heeft, is de zgn. puzzeltest, een test waarbij de testsubjecten gevraagd wordt om door elkaar gehaspelde woorden en constituenten weer in een voor het Nederlands aanvaardbare volgorde te zetten. Haeseryn (1990: 142) ging ervan uit dat de keuze voor de groene [PART+AUX] of de rode volgorde [AUX+PART] die de testsubjecten in dergelijke omstandigheden maken, niet beïnvloed kan zijn door de volgorde waarin de verschillende elementen aangeboden worden, omdat de elementen van de werkwoordelijke eindgroep niet in elkaars onmiddellijke buurt staan en de testsubjecten met andere woorden gedwongen worden om zelf een keuze te maken. In paragraaf 3.2.1 twijfelden we eraan of die assumptie wel vol te houden is, omdat altijd een van de werkwoordelijke elementen als eerste (i.e. het meest links in de zin) aangeboden moet worden en dat het daarom niet uit te sluiten is dat de uiteindelijke volgorde die gekozen wordt afhankelijk is van de relatieve positie van de werkwoordelijke elementen ten opzichte van elkaar.

Een indirecte bevestiging van die stelling vonden we in de resultaten van een eigen puzzelexperiment dat in oktober 2002 aan de K.U.Leuven afgenomen werd bij 136 tweedejaarsstudenten (*Taal- en letterkunde: Romaanse talen en Germaanse talen*). Het experiment was oorspronkelijk geconcipeerd om na te gaan (i) hoe vaak de testsubjecten voor de groene [PART+AUX] of rode volgorde [AUX+PART] kiezen en (ii) of die keuze geassocieerd is met de aan- vs. afwezigheid van een laatste zinsplaats (gedifferentieerd naar verschillende types—complement vs. adjunct—en verschillende lengtes—0 vs. 3 vs. 6 woorden). Vooraf werd bij de docenten Nederlandse taalkunde

nagevraagd of de betreffende studenten al in contact gekomen waren met het type woordvolgordevariatie dat hier bestudeerd wordt; dat bleek niet het geval te zijn. De testsubjecten kregen een formulier met 45 experimentele items voorgeschoteld, waarvan 20 fillers. Elk item begon met dezelfde matrixzin (*Het is een feit dat*) en werd gecontroleerd voor woordlengte en woordfrequentie (a.h.v. Celex; Baayen et al. 1993). Naast de experimentele items bevatte het formulier ook nog een instructietekst (die bij het begin van het experiment hardop werd voorgelezen) en een gedeelte waar regionale en sociale achtergrondinformatie over het testsubject gevraagd werd. Het invullen van het formulier nam tussen de 25 en de 35 minuten in beslag (appendix G).

De resultaten van het experiment gaven aan dat er geen effect was van de aanwezigheid van een laatste zinsplaats en ook de verschillen tussen de types en de lengte van de laatste zinsplaats hadden geen effect op de keuze voor de woordvolgorde, maar wel de wijze waarop we het materiaal hadden aangeboden:

Tabel 6.4: Absolute en relatieve proporties van de groene [PART+AUX] en rode [AUX+PART] woordvolgorde per type volgorde die in het puzzelexperiment aangeboden werd.

	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Aangeboden: PART eerst	72.51 % (1282/1786)	27.49 % (486/1768)
Aangeboden: AUX eerst	49.88 % (814/1632)	50.22 % (818/1632)

In tabel 6.4 wordt de keuze voor de groene [PART+AUX] of de rode volgorde [AUX+PART] bekeken als een functie van de aangeboden woordvolgorde ([PART] eerst of [AUX] eerst, gevolgd door eventueel tussenliggend materiaal, gevolgd door respectievelijk het hulpwerkwoord en het deelwoord). De verdeling van de gegevens in tabel 6.4 laten duidelijk zien dat de keuze voor de woordvolgorde in hevige mate bepaald wordt door de aangeboden woordvolgorde: biedt men eerst [PART] aan, dan kiezen de testsubjecten in verhouding vaker voor de groene volgorde [PART+AUX], biedt men eerst [AUX] aan, dan leidt dat tot een geïntensifieerde voorkeur voor de rode volgorde [AUX+PART]. Dat verschil is statistisch significant ($\chi^2 = 183.88$, $df = 1$, $p < .0001$) en de o.r. van 2.65 (B.I. = 2.29-3.07) maakt duidelijk dat de kans op een

groene volgorde [PART+AUX] meer dan 2.5 keer groter is dan een rode volgorde [AUX+PART] indien [PART] eerst aangeboden wordt (in vergelijking met een testzin waar [AUX] eerst aangeboden wordt).¹⁰ Omgekeerd zijn de odds voor de groene volgorde 2.63 keer kleiner dan de odds voor de rode volgorde [AUX+PART] als [AUX] eerst aangeboden wordt (o.r. = .38, B.I. = 0.33-0.44).

Die resultaten bevestigen met andere woorden de hypothese dat in een puzzelexperiment als dat van Haeseryn (1990) een zekere vorm van persistentie werkzaam geweest kan zijn, waardoor testsubjecten het verbale element dat het eerst wordt aangeboden vooraan in de werkwoordelijke eindgroep zetten. Hoe dat persistentie-effect precies begrepen moet worden—psycholinguïstisch of puur mechanisch¹¹, zoals in paragraaf 3.2.1 gesuggereerd werd—zal toekomstig onderzoek moeten uitwijzen. In ieder geval maken de resultaten van de puzzeltaak duidelijk dat niet alle geobserveerde persistentie-achtige effecten herleid kunnen worden tot een (mogelijk) effect van idiolectische stilistische voorkeur.

6.3 De impact van de frequentie van het deelwoord

In deze paragraaf komt het effect van deelwoordfrequentie aan bod. Aangezien de aard en de sterkte van een dergelijk effect op de woordvolgorde in de tweeledige werkwoordelijke eindgroep nog nooit bestudeerd is, kunnen we volstaan met een erg beknopte status quaestionis en een veeleer algemene hypothese (respectievelijk paragraaf 6.3.1 en 6.3.2). In paragraaf 6.3.3 presenteren en evalueren we de resultaten van ons onderzoek (paragraaf 6.3.3).

6.3.1 *Status quaestionis*

Ondanks het feit dat de impact van de frequentie van het deelwoord op de keuze voor de groene [PART+AUX] of rode [AUX+PART] volgorde nog nooit bestudeerd is, ligt de selectie van de factor *frequentie* erg voor de hand, niet alleen vanwege de nadruk die erop gelegd wordt door een aantal theoretische modellen zoals *emergent grammar* en *usage-based grammar* (zie paragraaf 1.4) maar vooral omdat uit tal van experimenteel onderzoek reeds gebleken is dat de frequentie van een item een niet te miskennen invloed uitoefent op de taalproductie en -verwerking—“Higher frequency

of a unit or pattern results in a greater degree of what Langacker terms entrenchment, i.e. cognitive routinization, which affects the processing of a unit” (Kemmer & Barlow 2000: x; zie bv. ook Fenk-Oczlon 2001)—en van daaruit op het resultaat van die taalprocessen.

6.3.2 Hypotheses

Door de ontstentenis aan onderzoek naar de frequentie van het deelwoord, zullen we de veeleer algemene hypothese [HYP 6.3] testen:

[HYP 6.3] De frequentie van het deelwoord heeft een invloed op de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde (paragraaf 6.3.3).

6.3.3 Resultaten en discussie

Voor de studie naar het effect van deelwoordfrequentie hebben we gebruik gemaakt van de lemmafrequentie-informatie in Celex (Baayen et al. 1993), d.w.z. dat we bijvoorbeeld voor het deelwoord *gekregen* niet de frequentie van het token *gekregen* opgezocht hebben, maar de frequentie van het lemma *krijgen* (met de frequentie van alle corresponderende werkwoordsvormen, zoals pv's, infinitieven en deelwoorden, erbij; verder: lemmafrequentie van het deelwoord). Dat de voorkeur gegeven werd aan een Celex-frequentielijst boven een frequentielijst die onttrokken is aan het CONDIV-corpus zelf, heeft, net als in hoofdstuk 5, enkel een praktische reden: Celex is gelemmatiseerd en CONDIV niet, waardoor de opzoeking van de lemmafrequenties per deelwoord veel sneller verloopt. Van een aantal deelwoorden werd noch token-, noch lemma-informatie gevonden. Omdat het daarbij vaak om samengestelde deelwoorden ging, werd daarom de frequentie van het minder gelede deelwoord genomen. In concreto gaat het om *binnengesijpeld*, *doordesemd*, *hergebruikt*, *heropgebouwd*, *overgesneden*, *uitgesmokkeld*, *vooruitgeschoven*, *weggeplukt* en *weggezuiverd*. Ondanks die procedure werd van een aantal deelwoorden geen frequentie-informatie gevonden. Het gaat daarbij om *gedepolitiseerd*, *gedynamiteerd*, *gemarginaliseerd*, *ontmijnd*, *verstrengd* en *weggedeemsterd*.

Het spreekt voor zich dat het te veel ruimte zou vergen om voor elk deelwoord de lemmafrequentie op te sommen, maar in tabel 6.5 hebben we alvast de top-20 van de frequentste werkwoorden (volgens lemmafrequentie) opgenomen:

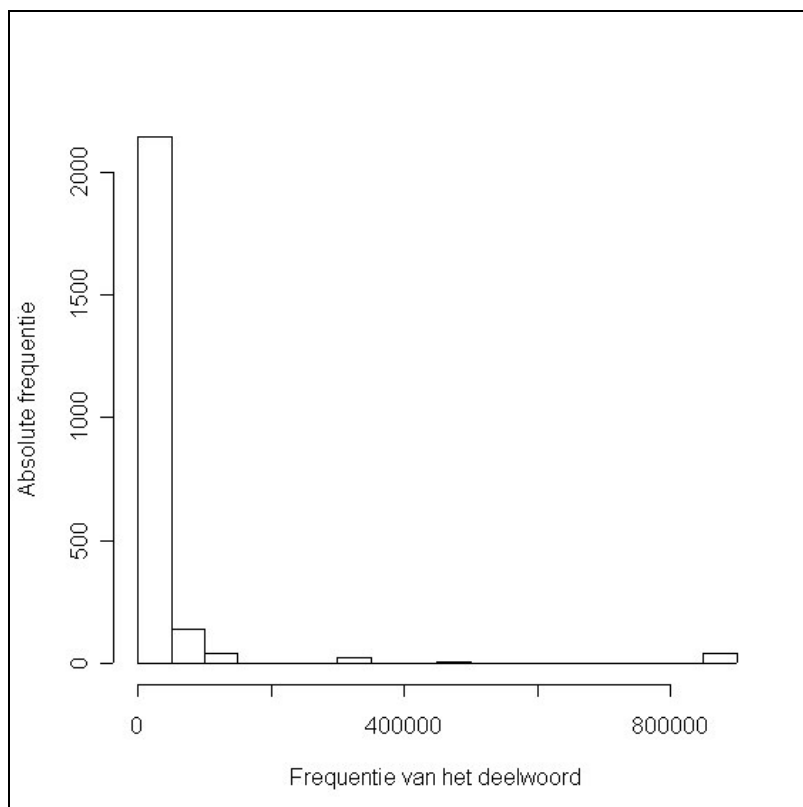
Tabel 6.5: Top-20 van de frequentste werkwoorden in de dataset.

Deelwoord (infinitief)	Frequentie
Geweest (zijn)	852027
Gehad (hebben)	474965
Geworden (worden)	321883
Gekund (kunnen)	209812
Gegaan (gaan)	137589
Gekomen (komen)	135035
Gezegd (zeggen)	134802
Gezien (zien)	98375
Gemaakt (maken)	86305
Gestaan (staan)	79616
Gedaan (doen)	71277
Gelaten (laten)	66787
Geweten (weten)	66735
Gedacht (denken)	58538
Gegeven (geven)	54472
Gebleven (blijven)	52837
Gevonden (vinden)	51631
Gehouden (houden)	46508
Gekregen (krijgen)	45929
Genomen (nemen)	43709

Niet geheel onverwacht kan uit tabel 6.5 opgemaakt worden dat de meest frequente werkwoorden in de Celex-database de werkwoorden zijn die vaak als hulpwerkwoord gebruikt worden (*zijn*, *hebben* en *worden*). Verder vallen de immense frequentieverschillen in de hoogste regionen van die top-20 op: zo komt het frequentste deelwoord (*geweest*, $n = 852027$) bijna twee keer zo vaak voor als het tweede frequentste deelwoord (*gehad*, $n = 474965$) en hebben slechts 7 van de 20 deelwoorden een lemmafrequentie die hoger is dan 100000. Op basis van die eerste vaststellingen kan al vermoed worden dat, net als bij de continue variabelen in hoofdstuk 4 en 5, de continue variabele *lemmafrequentie van het deelwoord* eveneens

niet normaal verdeeld zal zijn. Om ons daar een visuele voorstelling van te maken, hebben we in figuur 6.2 de lemmafrequentie van de verschillende deelwoorden (gebaseerd op Celex) uitgezet ten opzichte van de frequentie waarmee die deelwoorden in onze dataset voorkomen:

Figuur 6.2: Verdeling van de lemmafrequentie van het deelwoord ten opzichte van de frequentie van de deelwoorden in de dataset.



Het hoeft geen betoog dat de verdeling in figuur 6.2 erg ongelijkmatig is, daar ze een erg duidelijke bias in de richting van de minst (lemma)frequentie deelwoorden vertoont, hetgeen wil zeggen dat er heel veel deelwoordvormen in onze dataset voorkomen die een veeleer lage lemmafrequentie hebben. Ter illustratie van die ongelijkheid kan vermeld worden dat maar liefst 95.44% van de deelwoorden in de dataset een lemmafrequentie heeft die lager is dan 100000 ($n = 2281$)¹² en 74.56% van de deelwoorden een lemmafrequentie heeft die lager is dan 10000 ($n = 1782$). De scheefheidsmaat die in hoofdstuk 4 voor het eerst gebruikt is, bevestigt die vaststelling—de scheefheidswaarde voor de verdeling van de deelwoorden volgens frequentie is 6.22 (een perfect symmetrische verdeling heeft een scheefheidswaarde van 0)—en de statistische verificatie voor de niet-normale verdeling wordt geleverd

door de Shapiro-Wilktest—voor de volledige dataset: $W = 0.2522$, $p < .0001$; voor de deelwoorden in de groene volgorde [PART+AUX]: $W = 0.1598$, $p < .0001$; voor de deelwoorden in de rode volgorde [AUX+PART]: $W = 0.2913$, $p < .0001$.

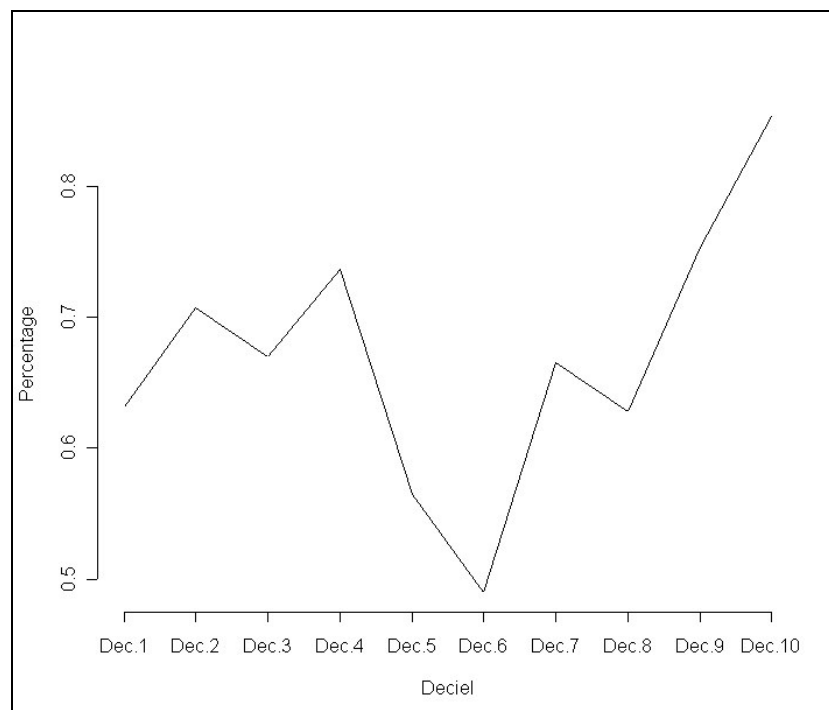
Daar de assumptie van normaliteit niet volgehouden kan worden, moeten we opnieuw onze toevlucht zoeken bij de Wilcoxon rank sum test om na te gaan of de lemmafrequentie van het deelwoord een invloed heeft op de woordvolgordeplaatsing. We herhalen kort dat de Wilcoxontest de deelwoorden sorteert volgens Celex-lemmafrequentie en vervolgens aan elk deelwoord een rangnummer toekent dat de positie van het deelwoord in het continuüm weergeeft (het minst frequente deelwoord krijgt het rangnummer 1 mee). Als er geen associatie is tussen de lemmafrequentie van het deelwoord en de woordvolgorde, dan kan men verwachten dat het gemiddelde rangnummer voor de deelwoorden in de groene volgorde [PART+AUX] niet verschilt van het gemiddelde rangnummer in de rode volgorde [AUX+PART] (dat is de nulhypothese van de Wilcoxontest). Indien zou blijken dat het gemiddelde van de rangnummers voor een van de volgordevarianten toch statistisch significant groter of kleiner is dan de andere variant, dan kan aan de hand daarvan besloten worden dat er een verband is tussen de positie van de deelwoorden in het continuüm en de keuze voor een van de volgordes. Het resultaat van de Wilcoxontest geeft te kennen dat de nulhypothese verworpen moet worden ($W = 575482$, $p < 0.0005$) en dat er met andere woorden inderdaad een verband bestaat tussen de lemmafrequentie van het deelwoord en de keuze voor een van de volgordevarianten (bevestiging [HYP 6.3]). In welke richting het effect van de deelwoordfrequentie gaat, kan afgeleid worden uit de centrummaten die in tabel 6.6 gepresenteerd worden.

Tabel 6.6: Centrumgegevens van de lemmafrequentie van het deelwoord voor de groene [PART+AUX] en de rode [AUX+PART] volgorde.

	Frequentie van het deelwoord Groene volgorde	Frequentie van het deelwoord Rode volgorde
Gemiddelde rangnummers	460.48	481.30
Mediaan frequentie	1903.5	2027
Gemiddelde frequentie	13064.43	38182.86

Tabel 6.6 laat zien dat zowel voor het gemiddelde rangnummer, de mediaan als voor het gemiddelde de deelwoorden in de rode volgorde [AUX+PART] hoger scoren dan de groene volgorde [PART+AUX]. Dat wijst er in alle gevallen op dat de deelwoorden in de rode volgorde [AUX+PART] gemiddeld genomen frequenter zijn dan de deelwoorden in de groene volgorde [PART+AUX]. Bij verder onderzoek blijkt er zelfs een (lichte) positieve correlatie te bestaan tussen de proportie rode volgordes [AUX+PART] en de lemmafrequentie van het deelwoord (in decielen; $r = 0.05$, $ASE = 0.02$). In figuur 6.3 wordt het verband tussen de lemmafrequentie van de deelwoorden in onze dataset in decielen—i.e. we hebben de deelwoorden in onze dataset volgens lemmafrequentie gerangschikt en vervolgens in 10 gelijke stukken verdeeld (elk stuk bevat $2390/10 = 239$ deelwoorden)—en de relatieve proportie rode woordvolgordes [AUX+PART] visueel weergegeven.

Figuur 6.3: Verdeling van de lemma frequentie van het deelwoord in decielen ten opzichte van de relatieve frequentie van de rode volgorde [AUX+PART].



Met uitzondering van het middelste interval (deciel 4 t.e.m. 7) blijkt er inderdaad een graduele toename te zijn van het aantal rode woordvolgordes [AUX+PART] naarmate de lemmafrequentie van het deelwoord stijgt. Hoe het effect van de lemmafrequentie van het deelwoord precies begrepen moet worden, is echter vooralsnog niet duidelijk. In ieder geval lijkt het erop dat geen van de vorige taalinterne verklaringsprincipes in aanmerking komt om het effect een adequate linguïstische interpretatie te geven. Indien dat wel het geval zou zijn, dan zou aangenomen moeten worden dat (i) hoogfrequente deelwoorden een klemtoon aan het begin van het woord hebben (zoals scheidbaar samengestelde werkwoorden) en laagfrequente woorden een klemtoon in het midden of aan het eind (volgens het accentueringsprincipe) en/of dat (ii) hoogfrequente deelwoorden minder adjectivisch zijn of een werkingsgerichtere semantiek hebben (volgens het semantische principe). Aangezien er geen aanwijzingen zijn dat die veronderstellingen met de werkelijkheid overeenstemmen—geen van de deelwoorden uit de top-20 van de frequentste deelwoorden (tabel 6.5) heeft de klemtoon op de eerste syllabe en een heel aantal van die deelwoorden heeft een toestandsbetekenis—zullen we in het volgende hoofdstuk het effect van de lemmafrequentie van het deelwoord op een andere manier trachten te interpreteren.

6.5 Samenvatting

In dit hoofdstuk hebben we de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep met een deelwoord en een hulpwerkwoord vanuit een psycholinguïstisch perspectief bestudeerd. De analyses waarover in dit hoofdstuk gerapporteerd is, hebben daarbij duidelijk gemaakt dat de keuze voor de groene [PART+AUX] of rode [AUX+PART] volgorde beïnvloed wordt door de *frequentie van het deelwoord* en door wat in de psycholinguïstiek *syntactische persistentie* genoemd wordt, i.e. de neiging van taalgebruikers om een bepaald syntactisch patroon dat recent geactiveerd is te recycleren. We vatten voor de beide factoren de belangrijkste resultaten samen.

6.5.1 Syntactische persistentie

Om een antwoord te vinden op de vraag of de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde in de tweeledige werkwoordelijke eindgroep (targeteindgroep) bepaald wordt door de volgorde van het hulpwerkwoord en het deelwoord in een voorafgaande tweeledige werkwoordelijke eindgroep (pre-eindgroep), hebben we de voorafgaande context van de targeteindgroepen binnen een span van 10 zinnen afgespeurd. Daarbij stelden we in eerste instantie vast dat in bijna de helft van de gevallen een tweeledige pre-eindgroep aanwezig was. In tweede instantie bleek dat de volgorde van het deelwoord en het hulpwerkwoord in die pre-eindgroep erg bepalend was voor de volgorde in de targeteindgroep: indien een tweeledige pre-eindgroep in de rode volgorde [AUX+PART] voorafgaat, dan stijgt de keuze voor de rode volgorde [AUX+PART] in de targeteindgroep en, omgekeerd, indien een groene pre-eindgroep [PART+AUX] voorafgaat, dan wordt in de targeteindgroep vaker een groene volgorde [PART+AUX] verkozen. Uit die vaststellingen blijkt dat taalgebruikers de keuze van de woordvolgorde in de werkwoordelijke eindgroep laten afhangen van de volgorde die in een voorafgaande eindgroep gebruikt is of, psycholinguïstisch geformuleerd, dat taalgebruikers de (onbewuste) neiging hebben om de volgorde van een eindgroep te recycleren, omdat dat de computationele last die gepaard gaat met het taalproductieproces reduceert.

Nu moet bij die interpretatie onmiddellijk de opmerking gemaakt worden dat de gegevens in onze dataset niet als een direct bewijs voor het werkzaam-zijn van een

syntactisch persistentie-effect kunnen gelden (daarvoor hebben we een psycho-experimentele setting nodig). Een alternatieve interpretatie die hier niet uit te sluiten valt, is dat deze of gene redacteur een persoonlijke voorkeur heeft voor de groene [PART+AUX] of de rode [AUX+PART] volgorde en die volgorde dan ook bewust gebruikt in zijn of haar artikelen. Van een psycholinguïstisch persistentie-effect is dan uiteraard geen sprake meer, wel van een stilistisch persistentie-effect.

In een tweede stap hebben we de resultaten uit de eerste analyse trachten te verifiëren voor een uitgebreidere set aan pre-eindgroepen, met name twee- en meerledige werkwoordelijke eindgroepen die minstens een hulpwerkwoord en een deelwoord bevatten en een reële variatiemogelijkheid inhouden. De resultaten van de analyse gaven ook hier te kennen dat er een verband is tussen de gekozen volgorde in de pre-eindgroep en de volgorde in de targeteindgroep (de targeteindgroep werd overigens in bijna twee derde van de gevallen voorafgegaan door een twee- of meerledige pre-eindgroep). De interpretatie van dat verband werd echter gehinderd doordat de meerledige clusters uiteraard niet eenvoudig in termen van groene [PART+AUX] en rode [AUX+PART] woordvolgorde te vatten zijn. Met behulp van een aantal statistische tests kwamen we echter tot het inzicht dat de verschillende types pre-eindgroepen herleid konden worden tot twee basistypes: een basistype waarbij het hulpwerkwoord rechts staat van het van het eerste bepalende werkwoord (het rode basistype [AUX+PART]) en een basistype waarbij het hulpwerkwoord links staat van het van het eerste bepalende werkwoord (het groene basistype [PART+AUX]). Wanneer we dan vervolgens de bovenstaande analyse nog eens herhaalden voor de twee ‘hybride’ basistypes, dan stelden we opnieuw hetzelfde effect als in de eerste analyse vast.

Geconfronteerd met dat persistentie-effect hebben we ons in een derde stap afgevraagd of de afstand tussen de voorafgaande werkwoordelijke eindgroep en de targeteindgroep een rol speelt bij de kracht van het persistentie-effect. Immers, aangezien het persistentie-effect als een psycholinguïstisch fenomeen beschouwd kan worden en uit allerlei onderzoek gebleken is dat de taalgebruiker niet over een oneindige geheugencapaciteit beschikt, is het aannemelijk dat het persistentie-effect in kracht afneemt naarmate de afstand tussen de pre-eindgroep en de targeteindgroep groter wordt. Dat bleek echter niet het geval te zijn: uit onze analyses kwam naar

voren dat het persistentie-effect stabiel blijft over de verschillende afstanden heen (de maximale afstand tussen de targeteindgroep en de pre-eindgroep bedraagt 10 zinnen).

In een slotparagraaf, die veeleer methodologisch getint was, hebben we aandacht geschonken aan de heimelijke beïnvloeding van een persistentie-effect in experimentele designs. In hoofdstuk 3 hadden we nl. het vermoeden geuit dat de resultaten van Haeseryns (1990) puzzelexperiment beïnvloed zouden kunnen zijn door een dergelijk syntactisch-persistentie-effect (terwijl dat niet de bedoeling was). Aan de hand van een eigen puzzelexperiment hebben we kunnen aantonen dat het inderdaad niet uit te sluiten valt dat syntactische persistentie een invloed speelt in een psycholinguïstische puzzeltaak.

6.5.2 Frequentie van het deelwoord

Het onderzoek naar het effect van de frequentie vertrok vanuit de veeleer algemene idee dat de frequentie waarmee een taaleenheid (i.c. het deelwoord) gebruikt wordt onder meer de manier bepaalt waarop een taaleenheid afgebakend wordt in het geheugen, de mate waarin een taaleenheid toegankelijk is en de snelheid waarmee een taaleenheid verwerkt wordt in taalproductie- en taalperceptieprocessen. Daardoor is het niet ondenkbaar dat de frequentie van een taaleenheid, i.c. het deelwoord, ook een rol zou kunnen spelen bij de keuze tussen de rode [AUX+PART] en de groene volgorde [PART+AUX]. Uit de analyses bleek inderdaad dat de frequentie een invloed heeft op de woordvolgordevariatie, en wel in die zin dat deelwoorden die frequenter zijn vaker in de rode volgorde [AUX+PART] voorkomen dan deelwoorden die minder frequent zijn. Aangezien het effect van de factor nog nooit bestudeerd is, hebben we er voorlopig het raden naar waarom frequente deelwoorden vaker in de rode volgorde [AUX+PART] verschijnen. In hoofdstuk 7 komen we hierop terug.

Noten:

1. We herhalen hier dat een emergente benadering van taal gebaseerd is op de veronderstelling dat het strikte onderscheid tussen taalstructuur en taalgebruik onhoudbaar is en dat in plaats daarvan een voortdurende en dynamische wisselwerking tussen taalstructuur en taalgebruik aangenomen moet worden.

2. Daarbij moet wel opgemerkt worden dat het psycholinguïstische onderzoek van Hartsuiker & Westenberg (2000) met andere bedoelingen de woordvolgordeproblematiek benaderd heeft dan die waar wij primair in geïnteresseerd zijn. De woordvolgordevariatie in de werkwoordelijke eindgroep wordt daar nl. niet als het centrale studieobject beschouwd (zoals hier het geval is), maar als middel om de zgn. *linearisatiehypothese* (o.a. Hartsuiker, Kolk & Huiskamp 1999, Hartsuiker & Westenberg 2000) te verifiëren. De linearisatiehypothese is opgesteld als antwoord op de vraag hoe een spreker het voor elkaar krijgt om in een uiterst korte tijdspanne de geïntendeerde boodschap om te zetten in akoestisch percipieerbare woorden en zinnen—waarbij een aantal taken, zoals het selecteren van de woorden en de corresponderende fonologische informatie en het aaneenrijgen van de woorden conform de syntactische regels van de betreffende taal, razendsnel uitgevoerd moeten worden. De linearisatiehypothese zegt daarom dat er een mechanisme moet bestaan, het zgn. linearisatieproces, dat de woordvolgorde bepaalt van een reeds gedeeltelijk gespecificeerde hiërarchische structuur. In concreto komt het erop neer dat een zinsproductiemodel verondersteld wordt waarbij sprekers tijdens het spreken in een eerste, functionele fase de woorden met de semantische en syntactische info selecteren en syntactische functies toekennen aan die woorden. In een tweede, positionele fase, die gedeeltelijk beïnvloed wordt door de eerste fase, wordt de constituentenstructuur en de woordvolgorde bepaald. Het linearisatieproces veronderstelt nu dat de woordvolgorde pas na de constituentenstructuur bepaald wordt en niet gelijktijdig met de constituentenstructuur. Dat betekent met andere woorden dat de precieze woordvolgorde van constituenten en zinnen ondergespecificeerd blijft tijdens de bepaling van de functionele en hiërarchische relaties. Een dergelijk linearisatieproces kan worden gemotiveerd doordat het een geleidelijke verwerking toelaat: een spreker wacht niet tot wanneer de zin volledig grammaticaal geëncodeerd is alvorens met de articulatie te beginnen. Productie wordt vanuit dat perspectief verondersteld om op verschillende niveau's sequentieel gedeeltelijke stukken van zinnen te verwerken. Van zodra een niveau klaar is met een deel van de zin kan een ander niveau beginnen te werken op dat gedeelte (incrementele verwerking). Een dergelijke rolverdeling zorgt voor een efficiënte on line verwerking zodat een vloeiende output gegarandeerd kan worden. Het feit dat er, zoals dadelijk zal blijken, een syntactisch persistentie-effect vastgesteld kan worden bij de keuze tussen de groene [PART+AUX] of rode [AUX+PART] volgorde, interpreteren Hartsuiker & Westenberg als een bewijs dat de woordvolgordebepaling als een apart proces binnen het zinsproductiemodel beschouwd moet worden. In deze paragraaf willen we ons hoegenaamd niet mengen in de discussie (i) of er al dan niet een linearisatieproces verondersteld mag worden en (ii) of een syntactisch persistentie-effect al dan niet als een bevestiging van het linearisatieproces beschouwd mag worden. In plaats daarvan zullen we het persistentie-effect op een veel elementairder (psycholinguïstisch) niveau interpreteren, m.n. als bewijs dat geactiveerde syntactische structuren in het geheugen opgeslagen worden en (binnen een zekere tijdspanne) gerecycleerd kunnen worden bij de productie van een nieuwe zin.
3. Merk op dat we er hier van uitgaan dat het persistentie-effect niet gebonden is aan specifieke types bijzinnen en dat, met andere woorden, de persistentie van de woordvolgorde in de eindgroep van het ene type bijzin op het andere overgedragen kan worden.

4. Deelwoorden die synchroon als adjectief worden beschouwd (zoals *bestand* in *bestand zijn*) zijn niet in de analyse betrokken, omdat ze nauwelijks variatiemogelijkheid vertonen (zie hoofdstuk 6 voor de gebruikte criteria).
5. We laten de types die infrequent voorkomen vanaf hier buiten beschouwing.
6. Indien we de tweeledige clusters met een deelwoord en een hulpwerkwoord buiten beschouwing laten, dan zien we dat eveneens een statistisch significant effect bereikt wordt: $\chi^2 = 39.37$, $df = 6$, $p < .0001$.
7. Dat men hier ongeveer hetzelfde effect als in tabel 6.1 kan vaststellen, is uiteraard niet zo verwonderlijk, omdat al uit de gepartitioneerde chi-kwadraatanalyse gebleken was dat de ‘hybride’ categorieën in tabel 6.3 intern voldoende homogeen zijn en dus voldoende aansluiten bij de basiscategorieën uit tabel 6.1.
8. Merk op dat het hier gevallen betreft die binnen een context van 10 voorafgaande zinnen geen pre-eindgroep bevatten. Indien de voorafgaande context meer dan 10 zinnen bevat, is het met andere woorden mogelijk dat er zich verder terug wel degelijk een pre-eindgroep bevindt.
9. Indien men inzoomt op de volgordevoorkeur voor de gevallen waar een groene pre-eindgroep [PART+AUX] voorafgaat versus waar geen pre-eindgroep voorafgaat, dan kan men vaststellen dat de odds voor de groene volgorde [PART+AUX] gegeven een groene pre-eindgroep [PART+AUX] 53% hoger zijn dan de odds voor de groene volgorde [PART+AUX] gegeven geen pre-eindgroep (o.r. = 1.53, B.I. = 1.22-1.93).
10. Wanneer we de aangeboden volgordes nog verder opsplitsen in een categorie waarbij de werkwoordelijke elementen gescheiden zijn door ander materiaal en een categorie waarbij de verbale elementen niet van elkaar gescheiden zijn, dan zien we dat dat voor de testzinnen in de groene volgorde [PART+AUX] geen statistisch significant verschil oplevert ($\chi^2 = 2.01$, $df = 1$, $p > .05$), maar voor de rode volgorde [AUX+PART] wel ($\chi^2 = 26.51$, $df = 1$, $p < .0001$): de kans op een rode volgorde [AUX+PART] versus de kans op een groene volgorde [PART+AUX] stijgt met meer dan 1.5 keer indien de aangeboden testzin een niet-gesplitste rode volgorde [AUX+PART] bevat dan wanneer de werkwoordelijke elementen van elkaar gescheiden zijn (o.r. = 1.67, B.I. = 1.37-2.04).
11. Met *mechanisch* hebben we de situatie op het oog waarbij testpersonen de puzzeltaak uitvoeren via het principe van *trial and error*: testpersonen trachten eerst de eerst aangeboden constituent in de eerste slot te stoppen. Lukt dat niet, dan proberen ze het met de tweede aangeboden constituent, enzovoort, tot wanneer een van de aangeboden constituenten in de slot past. Vervolgens beginnen ze weer van vooraf aan met de eerste overgebleven constituent, enzovoort, tot wanneer een grammaticale Nederlandse zin verkregen wordt. Het spreekt voor zich dat een dergelijk persistentie-effect van een totaal andere aard is dan het ‘psycholinguïstische’ persistentie-effect.
12. Dat betekent dat de (volgens Celex) zeven frequentste deelwoorden (tabel 6.5) samen 109 keer voorkomen (= 2390-2281).

Hoofdstuk 7

Competitie en reïnterpretatie

7.1 Inleiding

In de vorige hoofdstukken hebben we een antwoord gezocht op de vraag welke mechanismen de keuze tussen de groene [PART+AUX] en rode [AUX+PART] woordvolgorde in het hedendaagse Nederlandse taalgebruik sturen. Aan de hand van reëel geproduceerde eindgroepen uit twee zorgvuldig gecompileerde corpora werd daartoe een hele reeks factoren geïdentificeerd, werd hun invloed op de woordvolgorde (statistisch) onderzocht en werden de resultaten geïnterpreteerd volgens een contextuele, prosodische, semantische of psycholinguïstische dimensie. Ondanks het feit dat het onderzoek zoals dat totnogtoe uitgevoerd is, bepaald evidente resultaten opgeleverd heeft, blijft het globale beeld toch ietwat diffuus. Zo is het voorlopig nog onduidelijk hoe de verschillende factoren zich tot elkaar verhouden in hun effect op de woordvolgorde, wat het gezamenlijke effect van de factoren op de woordvolgorde is en hoe de verschillende verklaringsdimensies met elkaar interageren.

Om daarop een beter zicht te krijgen, zullen we in dit slothoofdstuk de resultaten en conclusies uit de vorige hoofdstukken samenbrengen, tegenover elkaar afwegen en evalueren. In een eerste, statistische stap zullen we aan de hand van twee multivariate statistische technieken, een classificatieboomanalyse (CBA) en een logistische-regressieanalyse (LRA), de relaties tussen de verschillende factoren bestuderen door enerzijds na te gaan wat de relatieve impact is van elk van de factoren, en anderzijds eventuele interactie- of redundantieverbanden op te sporen¹.

Daarnaast zal het collectieve effect—het collectieve effect kan nl. meer zijn dan de simpele optelsom van de verschillende individuele effecten—en de predictieve kracht van de set van factoren gemeten worden (paragraaf 7.2).

Omdat het effect van de factoren uit de contextuele dimensie (*regio* en *register*) niet in dezelfde dataset onderzocht is als het effect van de factoren uit de taalinterne dimensies, kunnen we de dwarsverbanden tussen de taalinterne en taalexterne poot van dit onderzoek niet bestuderen en moeten we derhalve de globale statistische analyses beperken tot de taalinterne factoren (de verhouding tussen de taalexterne factoren onderling is in hoofdstuk 3 al onderzocht). In tabel 7.1 wordt een overzicht gegeven van alle taalinterne factoren (met hun respectievelijke waarden) waarvan in de vorige hoofdstukken vastgesteld is dat ze een statistisch significante invloed hebben op de woordvolgorde en die daarom gebruikt zullen worden in de CBA (7.2.1) en de LRA (7.2.2) (tussen haakjes bevindt zich het nummer van de paragraaf waar de resultaten van de bivariate analyses gepresenteerd en besproken zijn):

Tabel 7.1: Overzicht van de factoren die gebruikt zijn voor de classificatieboom- en logistische-regressieanalyse.

Factor	Waarden
Scheidbaarheid van het deelwoord (4.2.3.1)	{scheidbaar samengesteld, niet-scheidbaar samengesteld}
Definietheid / Informationaliteit van de laatste preverbale substantiefgroep / het laatste linkerwoord (4.3.3.1 en 4.3.3.2)	{indefiniet nomen, definitief nomen, andere} {hoog informatieel, intermediair informatieel, laag informatieel}
Inherentie van de laatste preverbale constituent (4.3.3.3)	{numerische waarde op basis van de n-gecontroleerde LLR ² } {significant bigram, niet-significant bigram} ³
Lengte van het middenstuk in termen van woorden (4.3.3.4)	{0-2, 3-5, 6-8, 9-11, 12-14, >14}
Aantal onbeklemtoonde lettergrepen tussen het deelwoordaccent en het accent van het laatste preverbale woord (4.3.3.5)	{0&1, 2&3, >3}
Laatste zinsplaats (4.4.3.1)	{aanwezigheid, afwezigheid}
Grammaticale relatie tussen laatste zinsplaats en hoofd (4.4.3.2)	{complement, adjunct}
Type hoofd van de laatste zinsplaats (4.4.3.2)	{verbaal, preverbaal}

Aantal onbeklemtoonde lettergrepen tussen het deelwoordaccent en het accent van het eerste postverbale inhoudswoordwoord (4.4.3.3)	{0&1, 2&3, >3}
Status van het deelwoord (5.2)	{numerische ongewogen waarde op basis van het algoritme} {numerische gewogen waarde op basis van het algoritme} ⁴
Type hulpwerkwoord (5.3)	{koppelwerkwoord, hulpwerkwoord van tijd, <i>zijn</i> als hulpwerkwoord van het passief, <i>worden</i> , onclassificeerbaar}
Aan- vs afwezigheid van een pre-eindgroep (6.2)	{groene volgorde [PART+AUX], rode volgorde [AUX+PART], geen pre-eindgroep} ⁵
Lemmafrequentie van het deelwoord (6.3)	numerisch

In een tweede stap zullen we de resultaten van de globale statistische analyses op een meer linguïstische manier trachten te interpreteren. Daarbij zal een inhoudelijke round-up en re-evaluatie van de verschillende interne taaldimensies ondernomen worden en zal een eerste poging gedaan worden om de verschillende dimensies te reorganiseren op basis van een nieuw geïntroduceerde verklaringsdimensie (paragraaf 7.3).

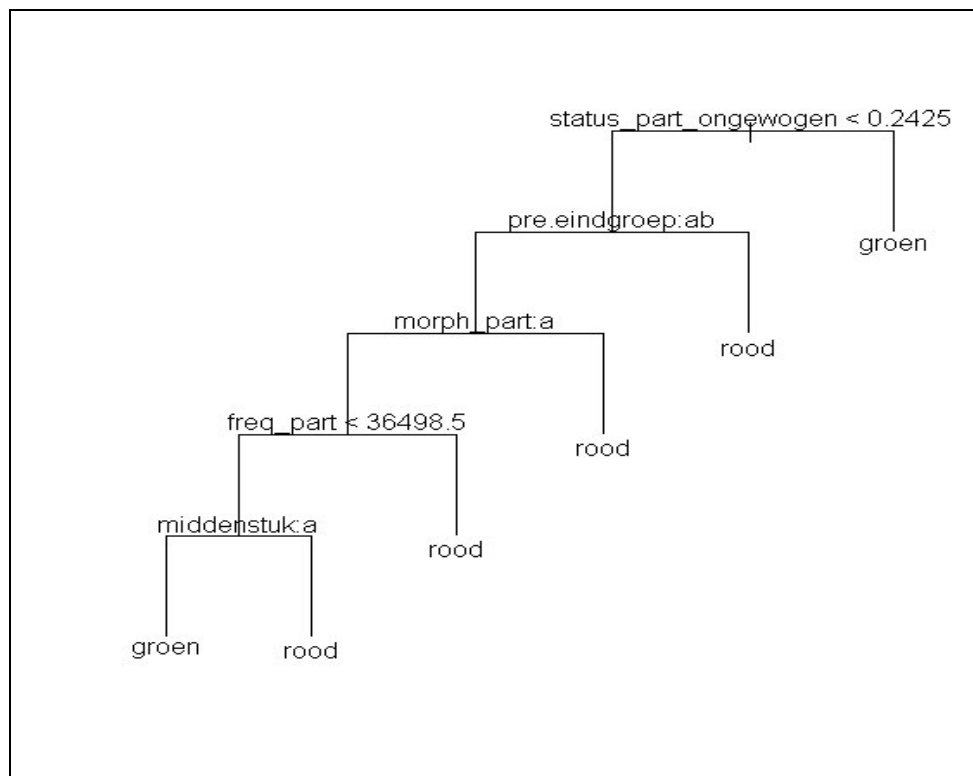
7.2 Multivariate statistische analyse⁶

7.2.1 Classificatieboomanalyse

In tegenstelling tot de logistische regressie die in de volgende paragraaf aan bod komt, behoort de CBA (zie bv. Harrell 2001: 26-27 en, toegepast op een linguïstische case study, Mendoza-Denton et al. 2003: 128-131) niet tot de inferentiële statistische technieken, maar is het een descriptieve, exploratieve, non-parametrische⁷, multivariate statistische techniek die een eerste inzicht kan verschaffen in de (hiërarchische) verhouding tussen de verschillende taalinterne predictorvariabelen en de graad van succes waarmee aan de hand van die variabelen de groene [PART+AUX] of de rode woordvolgorde [AUX+PART] voorspeld kan worden.

Om dat te realiseren wordt in een eerste stap de volledige dataset (i.c. $n = 2390$) op basis van een van de geselecteerde predictorvariabelen zo goed mogelijk opgesplitst in twee groepen, een groep met groene woordvolgordes [PART+AUX] en

een groep met rode woordvolgordes [AUX+PART]: een ideale splitsing minimaliseert de (woordvolgorde)variatie binnen een groep en maximaliseert de (woordvolgorde)variatie tussen de twee groepen.⁸ Als de meest ideale splitsing gevonden is (de specifieke techniek die wij verderop zullen gebruiken om de meest ideale splitsing op te sporen, is de maximale reductie van deviantie⁹), wordt voor elke resulterende knoop gekeken wat de interne variatie is. Indien geen of onvoldoende variatie meer gevonden kan worden, dan vormt die knoop een zgn. terminale knoop, als er wel nog variatie is, dan wordt die knoop opnieuw, maar nu op basis van een van de andere geselecteerde predictorvariabelen, opgesplitst in twee groepen, enzovoort, tot wanneer de resulterende knoop voldoende homogeen is of te weinig data bevat om verder op te splitsen. Het spreekt voor zich dat de CBA in principe zo lang zou kunnen doorgaan tot wanneer elke terminale knoop precies 1 observatie bevat, zodat er nergens nog interne variatie is (zgn. overclassificatie). Niettegenstaande dat een dergelijke classificatieboom perfect bij de onderliggende dataset (de basissample) zou passen, heeft hij nauwelijks generaliserende en predictieve waarde ten aanzien van een onafhankelijke dataset. Het is daarom zaak om de boom precies zo complex te laten worden dat hij de datapunten uit de basissample in rekening kan brengen en precies zo eenvoudig dat nieuwe datapunten erdoor voorspeld kunnen worden. Om de juiste grootte van de boom te bepalen, bestaan verschillende technieken (*tree pruning*). Aangezien in dit onderzoek enkel een (kleine) basissample voorhanden is en geen onafhankelijke testsample, ligt de keuze voor de zgn. *10-fold cross-validation* voor de hand. De basissample wordt daarbij in tien gelijke subsamples verdeeld, op basis van 9 subsamples wordt vervolgens een classificatieboom gegenereerd die tot slot getest wordt op de 10^{de} subsample. Die procedure wordt 9 keer herhaald, waarbij elke subsample 9 keer deel uitmaakt van de basissample en 1 keer als testsample gebruikt wordt. De resultaten van die *10-fold cross-validation* worden tot slot gemiddeld. Het resultaat van de CBA voor onze dataset wordt gepresenteerd in figuur 7.1:

Figuur 7.1: Classificatieboom.

Voor de datapunten in onze dataset blijkt uit figuur 7.1 dat de eerste en belangrijkste splitsing gerealiseerd wordt door de ongewogen variant van de predictorvariabele *status van het deelwoord* (continue variabele): de eindgroepen die een deelwoord bevatten met een adjectiviteitsgraad lager dan 0.2425 worden gesplitst van de eindgroepen die een deelwoord bevatten met een adjectiviteitsgraad hoger dan 0.2425 (merk op dat de adjectiviteitsgraad een waarde kan aannemen tussen 0 en 1, waarbij 0 staat voor extreem verbaal en 1 voor extreem adjectivisch). Voor die laatste groep, die een terminale knoop vormt in de classificatieboom, wordt telkens een groene woordvolgorde [PART+AUX] voorspeld. Om na te gaan wat de graad van succes van die voorspelling is, selecteren we in onze dataset alle eindgroepen met een deelwoord dat een adjectiviteitsgraad heeft die hoger ligt dan 0.2425. Daaruit blijkt dat die eindgroepen in 86% van de gevallen in de groene woordvolgorde [PART+AUX] voorkomen of, anders geformuleerd, dat de voorspelling tot een misclassificatiegraad van 14% leidt.

De niet-terminale knoop met eindgroepen die een deelwoord bevatten dat een adjectiviteitsgraad van minder dan 0.2425 heeft, wordt vervolgens verder opgesplitst in een knoop met eindgroepen die niet voorafgegaan worden door een pre-eindgroep

of voorafgegaan worden door een pre-eindgroep in de groene volgorde [PART+AUX] en een knoop met deelwoorden die voorafgegaan worden door een pre-eindgroep in de rode volgorde [AUX+PART]. Die laatste groep is een terminale knoop die een rode volgorde [AUX+PART] voorspelt. Die voorspelling leidt tot een misclassificatiegraad van 19%, of 81% van de deelwoorden in onze dataset die een adjectiviteitsgraad hebben die lager is dan 0.2425 en voorafgegaan worden door een rode pre-eindgroep [AUX+PART], komen in de rode volgorde [AUX+PART] voor.

Eindgroepen met een deelwoord dat een adjectiviteitsgraad van minder dan 0.2425 heeft en voorafgegaan wordt door een groene [PART+AUX] of door geen pre-eindgroep worden verder opgesplitst in een knoop waar het deelwoord niet-scheidbaar samengesteld is en een knoop waar het deelwoord wel scheidbaar samengesteld is. Die laatste groep constitueert opnieuw een terminale knoop (voorspelling: de rode volgorde [AUX+PART]) met een misclassificatiegraad van 19%.

De knoop met niet-scheidbare deelwoorden die voorafgegaan worden door geen of een groene pre-eindgroep [PART+AUX] en een adjectiviteitsgraad van minder dan 0.2425 hebben, wordt nog eens opgesplitst in een groep waarvan het deelwoord een frequentie heeft die groter is dan 36498.5 en een groep die een deelwoordfrequentie heeft van lager dan 36498.5. De knoop met de frequentste deelwoorden is een terminale knoop die de rode volgorde [AUX+PART] in 83% van de gevallen correct voorspelt (misclassificatiegraad: 17%).

De knoop met de minst frequente deelwoorden wordt nog een laatste keer opgesplitst in een knoop met zinnen die een middenstuk hebben van kleiner dan 3 woorden en een knoop met een middenstuk groter dan of gelijk aan 3 woorden. De beide knopen zijn nu terminaal: in de eerste knoop wordt de groene volgorde [PART+AUX] in 70% van de gevallen correct voorspeld (misclassificatiegraad: 30%), in de tweede knoop wordt de rode volgorde [AUX+PART] in 61% van de gevallen correct voorspeld (misclassificatiegraad: 39%).

In tabel 7.2 vatten we de resultaten van de CBA nog eens kort samen (met in de eerste kolom de predictorvariabele op basis waarvan de splitsing gerealiseerd wordt, in de tweede kolom het aantal overblijvende (nog te classificeren) observaties, in de derde kolom de resterende deviantie, in de vierde kolom de voorspelde

woordvolgorde, in de vijfde kolom de relatieve proportie van de voorspelde volgorde en in de zesde kolom de misclassificatiegraad; een asterisk duidt op een terminale knoop):

Tabel 7.2: Samenvattende statistieken van de classificatieboomanalyse.

splitsing	n	dev	voorspelling y	P(y)	misclassificatie
Volledige dataset	2390	3032	Rood	.66	.34
status_part_ongewogen < .2425	2149	2516	Rood	.72	.28
pre-eindgroep: groen, geen	1215	1555	Rood	.66	.34
morph_part: niet-scheidbaar	910	1214	Rood	.61	.39
freq_part < 36498	723	992	Rood	.55	.45
middenstuk < 3 woorden*	124	151	Groen	.70	.30*
middenstuk > 3 woorden*	599	800	Rood	.61	.39*
freq_part > 36894*	187	168	Rood	.83	.17*
morph_part = scheidbaar*	305	302	Rood	.81	.19*
pre-eindgroep: rood*	934	898	Rood	.81	.19*
status_part_ongewogen > .2425*	241	206	Groen	.86	.14*

Uit de samenvattende CBA-statistieken in tabel 7.2 kunnen we nu een aantal zaken concluderen.

In de eerste plaats blijkt dat we op basis van 5 predictorvariabelen in staat zijn om de keuze voor een van de woordvolgordes met een behoorlijke mate van succes te voorspellen. Als we in de laatste kolom van tabel 7.2 kijken naar de misclassificatiegraad in de terminale knopen, dan zien we dat, met uitzondering van de beide middenstukknopen, we in meer dan 80% van de gevallen correct de woordvolgorde konden voorspellen. Dat de misclassificatiegraad van de laatste terminale knopen (predictorvariabele *lengte van het middenstuk*) een stuk hoger ligt, wijst erop dat in vervolgonderzoek zeker nog gezocht zal moeten worden naar nieuwe predictorvariabelen die de homogeniteit in de knoop, bv. door nog een nieuwe splitsing te realiseren, kan verhogen. De globale misclassificatiegraad, over alle terminale knopen heen, bedraagt 23.93% (= 572 / 2390).

In de tweede plaats blijkt dat van alle bestudeerde predictorvariabelen de *status van het deelwoord*, *pre-eindgroep*, *scheidbaarheid van het deelwoord*, *frequentie van het deelwoord* en *lengte van het middenstuk* de belangrijkste zijn, i.e. het meest onderscheiden tussen eindgroepen met een groene volgorde [PART+AUX] en eindgroepen met een rode volgorde [AUX+PART]. Aangezien in vroeger wetenschappelijk werk de hiërarchische verhoudingen tussen de verschillende variabelen nooit onderwerp van (statistisch) onderzoek geweest zijn, was totnogtoe relatief onduidelijk welke variabelen de meeste impact hebben op de keuze voor de groene [PART+AUX] of de rode woordvolgorde [AUX+PART] en kunnen we onze resultaten dus ook niet vergelijken met een vroeger bepaalde maatstaf. Dat de status van het deelwoord tot dat selecte groepje van belangrijkste predictorvariabelen behoort, zal wellicht weinigen verbazen, gegeven de verbeterheid waarmee in de normatieve literatuur de rode volgorde [AUX+PART] verketterd wordt in eindgroepen met een adjectivisch deelwoord (een taalnorm die in zekere mate correspondeert met de tendensen in het taalgebruik). Het effect van de vier andere variabelen is in die zin veel minder intuïtief toegankelijk en daarom verrassender dan het effect van de status van het deelwoord. Het is in ieder geval opvallend dat alle taalinterne verklaringsdimensies door minstens een predictorvariabele vertegenwoordigd zijn—de semantische (1), prosodische (2) en psycholinguïstische (2) dimensie. Als men daarbij even aanneemt dat die variabelen inderdaad geïnterpreteerd kunnen worden tegen de achtergrond van de in de respectievelijke hoofdstukken voorgestelde verklaringsprincipes, dan kan daaruit voorlopig besloten worden dat de variatie in de werkwoordelijke eindgroep simultaan door verschillende types mechanismen beïnvloed wordt en dat er dus niet een mechanisme is dat het leeuwendeel van de variatie kan voorspellen (in paragraaf 7.3 komen we hierop terug).

In de derde en laatste plaats blijkt ook nog dat de ongewogen variant van de adjectiviteitsgraad in vergelijking met de gewogen variant en de variabele *type hulpwerkwoord* (die gedeeltelijk overlapt met de gewogen en ongewogen variabele) de meest optimale splitsing kan realiseren. Dat de andere operationaliseringën verderop in de boom niet meer aan bod komen, is makkelijk te begrijpen vanuit het feit dat de correlatie tussen die operationaliseringën zo groot is en de terminale knoop met groene woordvolgordes [PART+AUX] na de eerste splitsing zo zuiver is dat de andere operationaliseringën geen splitsingen meer kunnen realiseren. Andere opvallend

afwezige variabelen zijn de inherentie van het laatste preverbale woord (de LLR-meting uit hoofdstuk 4) en de enige echte prosodische variabele in deze studie (afstand tussen twee accenten in termen van onbeklemtoonde lettergrepen).

7.2.2 Logistische regressie¹⁰

De logistische regressie, die tot de groep van de zgn. *Generalized Linear Models* (GLM) behoort¹¹, is zonder enige twijfel de meest populaire techniek om een binaire responsvariabele (i.c. de groene volgorde [PART+AUX] of de rode volgorde [AUX+PART]) aan de hand van een stel predictorvariabelen statistisch te modelleren. In tegenstelling tot de CBA hierboven gebeurt dat niet door de dataset in twee groepen incrementeel te splitsen aan de hand van de predictorvariabele die de deviantie het meest reduceert, maar wordt aan elke predictorvariabele die in het logistisch-regressiemodel opgenomen is een regressiecoëfficiënt toegekend die de relatieve impact op de woordvolgorde, gecontroleerd voor alle andere predictorvariabelen, weergeeft. Alle coëfficiënten samen drukken de kans uit dat de woordvolgorde in de rode [AUX+PART] of in de groene woordvolgorde [PART+AUX] zal voorkomen, waarbij de grootte en de richting van de regressiecoëfficiënten de grootte van de kans bepalen. De relatie tussen de dichotome responsvariabele en de set van predictorvariabelen wordt over het algemeen uitgedrukt in de vorm van een regressievergelijking met aan de linkerkant van het isgelijktteken de responsvariabele, i.c. de woordvolgorde in de werkwoordelijke eindgroep, en aan de rechterzijde van het isgelijktteken de predictorvariabelen:

$$\log \left[\frac{P(Y=1)}{P(Y=0)} \right] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots$$

De responsvariabele (Y) wordt daarbij niet in termen van 100% voorspelde volgorde (zoals in de CBA), noch in termen van gemiddelden (zoals in een lineaire regressie) gemodelleerd, maar in termen van de log van de odds, i.e. de log van de kans dat Y = 1 (bv. de rode volgorde [AUX+PART]), gedeeld door de kans dat Y = 0 (bv. de groene volgorde [PART+AUX]). De reden daarvoor is dat het verband tussen de responsvariabele (als kans uitgedrukt) en de predictorvariabelen niet-lineair is, de voorspelde proportie een cijfer tussen 0 en 1 moet zijn om geïnterpreteerd te kunnen

worden—de responsvariabele is nl. binair, i.e. 0 versus 1—de responsvariabele niet normaal verdeeld is en de foutenterm geen constante variantie vertoont (zgn. heteroscedasticiteit; zie Rietveld & van Hout 1993: 327-361 en Agresti 1996: 74-79, 103-199 voor bijkomende informatie over die technische aangelegenheden). De symbolen α , β en X staan respectievelijk voor de constante, de regressiecoëfficiënten en de namen van de predictorvariabelen. Aangezien we slechts een steekproef uit de grotere populatie onderzoeken, moeten de regressiecoëfficiënten geschat worden. Dat gebeurt met de *maximum likelihood estimation* methode (zie Harrell 2001: 179-213, 228). Wanneer $\beta > 0$ dan stijgt de probabilmiteit op succes ($Y=1$) als de waarde van X stijgt, wanneer $\beta < 0$ dan daalt de probabilmiteit op succes als de waarde voor X stijgt, wanneer $\beta = 0$ dan is de probabilmiteit van succes voor alle waarden van X hetzelfde en zijn de respons- en predictorvariabele statistisch onafhankelijk van elkaar. De grootte van β bepaalt de mate waarin de probabilmiteit op succes stijgt, daalt of gelijk blijft per eenheid dat X stijgt en is met andere woorden een graadmeter voor de impact van de predictorvariabele op de woordvolgorde. Op basis van de regressiecoëfficiënt en de asymptotische standaardfout (ASE) kan dan ook de graad van significantie waarmee een predictorvariabele een effect heeft op de responsvariabele bepaald worden ($z = \beta / ASE$).

Laten we in een eerste stap een logistisch-regressiemodel, dat bestaat uit de vijf predictorvariabelen die in de CBA als de belangrijkste naar voren kwamen, op onze data fitten om de wat abstracte inleiding te concretiseren. Het model, waarvoor de resultaten in tabel 7.3 samengevat worden, modelleert de kans op een rode volgorde [AUX+PART]:

Tabel 7.3: Resultaat van de beknopte logistische-regressieanalyse.

Predictorvariabele	β	ASE	z-waarde	Pr (> z)	O.R.
Constante	-8.376e-01	1.479e-01	-5.665	1.47e-08 ***	
status deelwoord (ongew)	-4.654e+00	5.098e-01	-9.129	< 2e-16 ***	
frequentie deelwoord	3.853e-06	9.325e-07	4.133	3.59e-05 ***	
scheidbaar deelwoord	1.252e+00	1.319e-01	9.492	< 2e-16 ***	3.50
middenstuk: 3-5	8.748e-01	1.334e-01	6.560	5.40e-11 ***	2.40
middenstuk: 6-8	1.051e+00	1.478e-01	7.110	1.16e-12 ***	2.86
middenstuk: 9-11	1.185e+00	1.872e-01	6.331	2.43e-10 ***	3.27
middenstuk: 12-14	1.365e+00	2.907e-01	4.697	2.64e-06 ***	3.91
middenstuk: >14	9.896e-01	3.437e-01	2.879	0.00398 **	2.69
pre-eindgroep: geen	5.781e-01	1.272e-01	4.544	5.51e-06 ***	1.78
pre-eindgroep: rood	1.176e+00	1.293e-01	9.093	< 2e-16 ***	3.24

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

In de eerste kolom van tabel 7.3 zijn de verschillende predictorvariabelen opgenomen, in de tweede kolom vindt men de regressiecoëfficiënten, in de derde kolom de standaardfout, in de vierde kolom de z-waarde (berekend op basis van de waarden in de vorige twee kolommen) en in de laatste kolom staat de p-waarde (bepaald aan de hand van de z-waarde in de vorige kolom) die de graad van significantie weergeeft.

Laten we ons in eerste instantie even concentreren op de predictoren in de eerste kolom. Daaruit blijkt dat de categorische variabelen (scheidbaarheid van het deelwoord, lengte van het middenstuk en type pre-eindgroep) opgesplitst zijn volgens de waarden die de variabele kan aannemen, met uitzondering van één waarde. De reden daarvoor is dat nominale en ordinale variabelen in een regressie anders weergegeven en geëvalueerd worden dan continue variabelen. Categorische variabelen worden nl. gerepresenteerd door zgn. dummy variabelen (het aantal dummy variabelen = het aantal waarden dat de variabele kan aannemen), zodat voor elk van de waarden van de categorische variabele een regressiecoëfficiënt berekend kan worden. De regressiecoëfficiënt van een van de waarden, de referentiewaarde, wordt daarbij gelijkgeschakeld met 0 en fungeert daardoor als maatstaf om het effect van de andere waarden op de responsvariabele te evalueren (de procedure van de dummy codering voor categorische variabelen wordt uitgebreid toegelicht in Rietveld & van Hout 1993, Agresti 1996: 118-121 en Harrell 2001: 14). De referentiewaarden van de verschillende categorische variabelen worden traditioneel niet opgenomen in de

samenvattende statistieken (precies omdat ze op zich geen informatie dragen: $\beta = 0$). In ons geval gaat het om niet-scheidbaar deelwoord, middenstuk: 0-2 woorden, pre-eindgroep: groen.¹²

Wanneer we vervolgens onze aandacht richten op de regressiecoëfficiënten in de tweede kolom, dan kan in de eerste plaats uit *de richting* van de regressiecoëfficiënten afgeleid worden welke predictoren de kans op een rode volgorde [AUX+PART] doen toenemen en welke de kans op een rode volgorde [AUX+PART] doen afnemen: enkel de predictor *status van het deelwoord* blijkt de kans op een rode volgorde [AUX+PART] te doen afnemen (-4.65), de vier andere predictoren *frequentie van het deelwoord*, *lengte van het middenstuk*, *morfologische bouw van het deelwoord* en *type pre-eindgroep* stimuleren de kans op een rode volgorde [AUX+PART]. Uit de grootte van de regressiecoëfficiënten kan men in de tweede plaats ook afleiden *hoe sterk* de impact op de woordvolgorde is (hoe groter de afstand tot 0, hoe zwaarder de impact). Daaruit blijkt dat de predictor *status van het deelwoord* de grootste impact heeft op de keuze voor een van de woordvolgordes ($\beta = -4.65$), gevolgd door *lengte van het middenstuk: 12-14 woorden* ($\beta = 1.36$).

De regressiecoëfficiënten in de tweede kolom kunnen ook in termen van (conditionele) o.r.'s geïnterpreteerd worden, m.n. door de exponentiële functie e^β , waarbij e een constante is die bij benadering 2.718282 bedraagt—de exponentiële functie zorgt er nl. voor dat we de log aan de linkerkzijde van de regressievergelijking kwijt zijn. Op die manier kan bijvoorbeeld het effect van de predictor *scheidbaar samengesteld deelwoord*, die een regressiecoëfficiënt van 1.252 heeft, geïnterpreteerd worden als 'de kans op een rode volgorde [AUX+PART] versus de kans op een groene volgorde [PART+AUX] stijgt met 3.50 ($= e^{1.252}$) in het geval een eindgroep een scheidbaar samengesteld deelwoord bevat in vergelijking met eindgroepen die een niet-scheidbaar samengesteld deelwoord als hoofdwerkwoord hebben' (de referentiewaarde). In de laatste kolom van tabel 7.3 hebben we de verschillende o.r.'s opgenomen, met uitzondering van de continue variabelen, waarop het concept o.r. niet van toepassing is.

Op basis van de regressiecoëfficiënten in de tweede kolom en de standaardfouten in de derde kolom kan tot slot nog een z-waarde (vierde kolom) berekend worden, aan de hand waarvan we de significantie van elke predictor kunnen

evalueren ($z = \beta / \text{ASE}$; vijfde kolom). Uit tabel 7.3 blijkt dat alle predictorvariabelen een significante impact hebben op de woordvolgorde (op dezelfde manier als in de vorige hoofdstukken).

Naast de evaluatie van de individuele predictoren kan men op basis van een LRA ook nagaan hoe succesvol het regressiemodel in het algemeen is, of, anders geformuleerd, hoeveel van de geobserveerde variatie het model met de 5 predictoren kan verklaren en voorspellen. Vooraleer de verklarende en voorspellende waarde van het model te berekenen, moeten we echter als voorbereidende stap eerst nog even nagaan of het model wel geschikt is voor onze dataset, i.e. of het model in een goede fit van de data voorziet. Met behulp van een Pearson chi-kwadraattest vergelijken we daartoe het regressiemodel dat uit 5 predictorvariabelen bestaat met een zgn. gesatureerd model, i.e. een model dat evenveel parameters heeft als er observaties zijn en daardoor een exacte beschrijving geeft van de gegevens in de dataset. Indien het verschil tussen het spaarzame model met 5 predictorvariabelen en het gesatureerde model gering is en de p-waarde van de chi-kwadraatanalyse daardoor niet-significant is, dan kunnen we daaruit besluiten dat het spaarzame model de gegevens voldoende goed benadert (een spaarzaam model heeft altijd de voorkeur op het gesatureerde model, omdat het veel makkelijker interpreteerbaar is). Indien dat niet het geval is, dan is dat een teken dat er nog iets ontbreekt aan het spaarzame model. De resultaten van de chi-kwadraattest geven voor ons model aan dat de data niet goed gefit worden: het verschil tussen het gesatureerde en het spaarzame model is significant ($p < .000026$).

Een slecht fittend model als het onze kan op twee manieren aangevuld worden: enerzijds door extra predictorvariabelen mee in het model op te nemen, anderzijds door interactietermen op te nemen. Omdat interacties veel minder makkelijk interpreteerbaar zijn dan de hoofdeffecten (zeker als het om drie- of meerwegsinteracties gaat) en omdat interacties de kans op een multicollineariteitsprobleem doen toenemen, zullen we ons in deze studie voorlopig beperken tot de opname van additionele predictorvariabelen, meer bepaald alle andere predictorvariabelen uit tabel 7.1 die nog niet in het model zitten. Multicollineariteit duidt op het bestaan van hoge correlaties tussen de verschillende predictorvariabelen in een regressievergelijking. Die correlaties kunnen als gevolg hebben dat de regressiecoëfficiënten onbetrouwbaar en onstabiel zijn, waardoor het logistisch-

regressiemodel oninterpreteerbaar wordt. Multicollineariteit is overigens niet enkel een probleem voor modellen met interactietermen, ook modellen met enkel hoofdeffecten kunnen te lijden hebben onder multicollineariteit. Het spreekt voor zich dat we niet voor alle variabelen in ons onderzoek op voorhand kunnen weten of ze al dan niet zullen correleren, maar van een aantal variabelen ligt de correlatie voor de hand, m.n. voor de variabelen *status van het hulpwerkwoord/deelwoord* en voor de binaire (al dan niet n-gecontroleerde) en continue classificatie van de *inherentie van het laatste preverbale woord* (zie tabel 7.1 voor een overzicht), zodat het beter is om één vertegenwoordiger voor die clusters van variabelen te kiezen.

Om de keuze tussen de verschillende operationaliseringen niet op subjectieve gronden te laten plaatsvinden, vergelijken we de residuele deviantie van de verschillende logistische-regressiemodellen (de residuele deviantie is een maat voor verklaarde variantie: hoe minder residuele deviantie overblijft na het fitten van een model, hoe groter de verklarende kracht van het model). Zo wordt de residuele deviantie na het fitten van een model dat bestaat uit de predictorvariabele *ongewogen adjectiviteitsmeting* vergeleken met de residuele deviantie van een model dat bestaat uit de *gewogen adjectiviteitsmeting* en een model dat bestaat uit de variabele *status van het hulpwerkwoord*. Uit de deviantieanalyse blijkt dat de reductie in deviantie groter is voor de variabele *status van het hulpwerkwoord* (residu = 2657, df = 2385) dan voor de *ongewogen* (residu = 2769, df = 2388) en *gewogen adjectiviteitsmeting* (residu = 2730, df = 2388). Dezelfde procedure wordt herhaald om uit te maken welke van de variabelen *inherentie van het laatste preverbale woord* de grootste verklarende kracht heeft. Dat blijkt de binaire, n-gecontroleerde classificatie (residu = 2973, df = 2388) in significante en niet-significante bigrammen te zijn (binair, niet-gecontroleerde: residu = 3025, df = 2388; continue classificatie: residu = 3022, df = 2388).¹³

De resultaten van de uitgebreide logistische-regressieanalyse worden in tabel 7.4 voorgesteld (dus: alle predictorvariabelen uit tabel 7.1 zijn in het model opgenomen, behalve de net opgesomde):

Tabel 7.4: Resultaat van de uitgebreide logistische-regressieanalyse.

Predictorvariabele	β	ASE	z-waarde	Pr (> z)	O.R.
Constante	-3.689e+00	4.055e-01	-9.096	< 2e-16 ***	
hww van tijd	2.907e+00	2.337e-01	12.437	< 2e-16 ***	18.30
hww: passief zijn	2.057e+00	2.535e-01	8.113	4.94e-16 ***	7.82
hww: worden	2.462e+00	2.296e-01	10.722	< 2e-16 ***	11.73
hww: onclassificeerbaar	1.468e+00	3.389e-01	4.332	1.48e-05 ***	4.34
frequentie deelwoord	2.441e-06	7.738e-07	3.15	0.0016 **	
scheidbaar deelwoord	1.352e+00	1.821e-01	7.426	1.12e-13 ***	3.87
middenstuk: 3-5	7.083e-01	1.459e-01	4.856	1.20e-06 ***	2.03
middenstuk: 6-8	8.288e-01	1.613e-01	5.139	2.76e-07 ***	2.29
middenstuk: 9-11	8.270e-01	2.009e-01	4.116	3.85e-05 ***	2.29
middenstuk: 12-14	9.446e-01	3.001e-01	3.147	0.001649 **	2.57
middenstuk: >14	6.832e-01	3.653e-01	1.871	0.061407 .	1.98
pre-eindgroep: geen	5.404e-01	1.339e-01	4.035	5.47e-05 ***	1.72
pre-eindgroep: rood	1.190e+00	1.363e-01	8.730	< 2e-16 ***	3.28
informat.: intermediair	3.434e-01	2.051e-01	1.674	0.094125 .	1.41
informat.: hoog	6.625e-01	1.991e-01	3.327	0.000878 ***	1.94
significant bigram	8.163e-01	1.850e-01	4.412	1.02e-05 ***	2.26
lzp: complement van V	-7.423e-01	1.581e-01	-4.694	2.68e-06 ***	0.47
lzp: complement van N	1.947e-01	2.593e-01	0.751	0.452740	
#onbekl.syll.pre: 0&1	-1.543e-01	1.874e-01	-0.823	0.410371	
#onbekl.syll.pre: 2&3	1.060e-01	1.810e-01	0.586	0.558147	
#onbekl.syll.post: 0&1	4.485e-03	2.231e-01	0.020	0.983956	
#onbekl.syll.post: 2&3	-2.679e-02	1.863e-01	-0.144	0.885671	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

In tegenstelling tot het beknoptere model met 5 predictorvariabelen, fit het uitgebreide model met 9 predictorvariabelen de data in onze dataset wel goed ($p > .05$). Uit de resultaten die samengevat zijn in tabel 7.4 kunnen we een aantal zaken concluderen.

(i) De impact van de 5 variabelen die ook al in het beknopte model zaten, blijft ongewijzigd: de status van het deelwoord/hulpwerkwoord (referentiewaarde: hww: koppelwerkwoord), de frequentie van het deelwoord, de lengte van het middenstuk (referentiewaarde: middenstuk: 0-2 woorden), het type pre-eindgroep (referentiewaarde: pre-eindgroep: groen) en de morfologische bouw van het deelwoord (referentiewaarde: niet-scheidbaar deelwoord) behouden hun significante effect op de woordvolgorde.

(ii) Van de nieuwe predictorvariabelen blijken enkel de prosodische variabelen *afstand tussen het laatste preverbale accent en het deelwoordaccent* (referentiewaarde: #onbekl.syll.pre: >3 syllabes) en *afstand tussen het deelwoordaccent en het eerste postverbale accent* (referentiewaarde: #onbekl.syll.pre: >3 syllabes) de woordvolgorde niet te beïnvloeden. De predictorvariabelen *informationaliteit van het laatste preverbale woord* (referentiewaarde: informationaliteit: laag), *inherentie van het laatste preverbale woord* (referentiewaarde: niet-significant bigram) en *type laatste zinsplaats* (referentiewaarde: lzp: geen of adjunct) hebben wel een significante invloed op de keuze van de woordvolgorde, al moet daar wel bij opgemerkt worden dat dat voor een aantal individuele predictorwaarden van die variabelen niet geldt: *laatste zinsplaats dat als complement van een preverbaal nomen, adjectief of adverbium fungeert, intermediaire informationaliteit en lengte van het middenstuk dat meer dan 14 woorden in beslag neemt*.

(iii) De richting van de regressiecoëfficiënten geeft aan dat alle predictorvariabelen de kans op een rode volgorde [AUX+PART] bevorderen, behalve de predictorwaarde *laatste zinsplaats dat als complement van het werkwoord fungeert*. Niet onbelangrijk daarbij is dat de richting van die coëfficiënten in overeenstemming is met wat in de bivariate en conditionele analyses vastgesteld werd.

(iv) Aan de regressiecoëfficiënten kan men daarnaast ook de sterkte van het effect aflezen: hoe verder de regressiecoëfficiënt van 0 verwijderd is, hoe sterker het effect. Om een beter zicht te krijgen op de hiërarchische verhoudingen tussen de predictoren, hebben we de absolute waarde van de verschillende regressiecoëfficiënten gesorteerd van hoog naar laag (tussen haakjes hebben we de ook de o.r. opgenomen):

Hulpwerkwoord van tijd (o.r. = 18.30) > *hulpwerkwoord worden* (o.r. = 11.73) > *hulpwerkwoord zijn (passief)* (o.r. = 7.82) > *onclassificeerbaar deelwoord* (o.r. = 4.34) > *scheidbaar deelwoord* (o.r. = 3.87) > *pre-eindgroep in rode volgorde [AUX+PART]* (o.r. = 3.28) > *middenstuk: 12-14 woorden* (o.r. = 2.57) > *middenstuk: 6-8 woorden* (o.r. = 2.29) > *middenstuk: 9-11 woorden* (o.r. = 2.29) > *significant bigram* (o.r. = 2.26) > *laatste zinsplaats: complement van het werkwoord* (o.r. = 0.47) > *middenstuk: 3-5 woorden* (o.r. = 2.03) >

informaliteit: hoog (o.r. = 1.94) > geen pre-eindgroep (o.r. = 1.72) > frequentie van het deelwoord

Net als in de CBA blijkt de status van het deelwoord/hulpwerkwoord ook hier weer het meeste invloed uit te oefenen op de keuze voor de woordvolgorde, gevolgd door de morfologische bouw van het deelwoord en het type pre-eindgroep. Merk overigens op dat de variabele *deelwoordfrequentie*, waarvoor het concept o.r. niet van toepassing is omdat het een continue variabele is, op basis van de logistische regressie het minst sterk de woordvolgorde beïnvloedt, terwijl de deelwoordfrequentie in de CBA toch als vierde sterkste effect naar voren kwam. De resultaten van de logistische regressie zullen samen met de resultaten van de CBA in de volgende paragraaf linguïstisch geïnterpreteerd worden.

Zoals hierboven al opgemerkt werd, kan men naast de evaluatie van de individuele predictoren op basis van een LRA ook nagaan hoe succesvol het regressiemodel in het algemeen is, of, anders geformuleerd, hoeveel van de geobserveerde variatie het model met de 8 predictoren kan verklaren en voorspellen (hierboven zagen we al dat het uitgebreide model de data goed fit). De verklarende waarde van het model kan uitgedrukt worden in de totale reductie in deviantie ten opzichte van een model met enkel de constante (een zgn. dummy model): bedraagt de deviantie in het dummy model nog 3031.8, in het model met 8 predictoren is de deviantie met 706.1 significant gedaald tot 2325.7. De reductie in deviantie heeft geen kant-en-klare interpretatie en kan in die zin enkel gebruikt worden om concurrerende modellen met elkaar te vergelijken.

Stellen we ons tot slot nog de vraag wat de predictieve kracht is van het model met de 8 significante predictoren of, anders geformuleerd, in welke mate het model in staat is om de woordvolgorde te voorspellen aan de hand van de 8 significante predictoren uit de LRA. Een van de statistische maten die daarvoor gebruikt kan worden, is de concordantiemaat (c-maat). De c-maat geeft een index voor de correlatie tussen de voorspelde probabiliteiten en de geobserveerde waarden van de responsvariabele.¹⁴ De index kan een waarde aannemen tussen 0.5 (geen voorspellende kracht) en 1 (perfect voorspellende kracht). Voor het uitgebreide logistisch-regressiemodel dat we op onze data gefit hebben, levert dat een c-waarde van 0.803 op (na 100 bootstrapherhalingen¹⁵), waaruit blijkt dat het uitgebreide model

met 8 predictoren een zeer degelijk model is dat een niet onaanzienlijke predictieve kracht heeft: “A model having c greater than roughly .8 has some utility in predicting the responses of individual subjects” (Harrell 2001: 247).¹⁶

Desondanks moeten we om verschillende redenen beducht zijn voor een al te groot optimisme: “It is very typical to use the data to make decisions about the form of the model as well as about how predictors are represented in the model. Then, once a model is developed, the entire modeling process is routinely forgotten, and statistical quantities [...] are computed as if the resulting model were entirely prespecified” (Harrell 2001: 8).

In deze paragraaf hebben we het meest voor de hand liggende en meest eenvoudige logistische-regressiemodel op onze data gefit: we zijn uitgegaan van de verklarende variabelen die in de bivariate en trivariate analyses als significant naar voren kwamen en hebben via een eenvoudig additief model de woordvolgordevariatie in de werkwoordelijke eindgroep statistisch gemodelleerd. Extensies van het hier gepresenteerde model zijn in allerlei richtingen mogelijk én gewenst: een voor de hand liggende extensie is uiteraard de toevoeging van nieuwe predictorvariabelen met het oog op een nog betere verklaring en voorspelling van de variatie (waarbij niet uitgesloten is dat een of meerdere van de predictoren in ons model verwijderd moet worden) en de introductie van interacties tussen predictoren. Maar daarnaast zal ook het gamma aan statistische technieken verder uitgebuit moeten worden om vanuit statistisch oogpunt beter vat te krijgen op de variatie en de predictoren die de respons beïnvloeden (bv. via loglineaire analyses, clusteranalyses, correspondentieanalyses).

7.3 Globale linguïstische analyse

7.3.1 Inhoudelijke round-up en reïnterpretatie

Onze zoektocht naar het hoe en waarom van de werkwoordsvariatie in de tweeledige werkwoordelijke eindgroep heeft nu bijna zijn eindpunt bereikt. Nadat we in de vorige hoofdstukken al vastgesteld hadden dat een hele resem verklarende variabelen een belangrijke rol spelen bij de keuze tussen de groene [PART+AUX] en de rode volgorde [AUX+PART], hebben we in de vorige paragraaf aan de hand van een globale statistische analyse de verschillende individuele effecten van al die variabelen,

ingedeeld in drie dimensies, onderling met elkaar vergeleken (om technische redenen konden de taalexterne factoren uit hoofdstuk 3 niet in die vergelijking betrokken worden; zie paragraaf 7.1).

Uit die globale statistische analyse bleek dat we het fenomeen waarnaar we in hoofdstuk 1 nog verwezen als een type syntactische variatie dat bekend is voor zijn schijnbaar onontwarbaar kluwen van interagerende krachten, meer dan behoorlijk onder controle gekregen hebben: het leeuwendeel van de geobserveerde variatie in onze dataset kon nl. aan de hand van 5 à 8 variabelen verklaard en voorspeld worden (afhankelijk van het type statistische analyse).¹⁷ In tabel 7.5 sommen we de verschillende relevante variabelen per dimensie nog eens op:

Tabel 7.5: Overzicht van de statistisch relevante verklarende variabelen.

Prosodische dimensie	Semantische dimensie	Psycholinguïstische dimensie
Morfologische bouw van het deelwoord	Aard van het hulpwerkwoord/type deelwoord	Type pre-eindgroep
Informationaliteit van het laatste preverbale woord		Frequentie van het deelwoord
Inherentie		
Lengte van het middenstuk		
Type laatste zinsplaats		

Dat de verschillende variabelen in tabel 7.5 allemaal een significante bijdrage leveren tot de verklaring van de geobserveerde variatie, betekent niet dat ze die bijdrage allemaal in gelijke mate leveren. Zowel uit de CBA als uit de LRA bleek eenduidig dat de status van het deelwoord/hulpwerkwoord de belangrijkste invloed uitoefent op de keuze voor de groene [PART+AUX] of de rode volgorde [AUX+PART].

Verder bleken ook de morfologische bouw van het deelwoord (scheidbaar vs. niet-scheidbaar), de informationaliteit van het laatste preverbale woord, inherente preverbale woorden (i.c. werkwoordelijke vaste verbindingen), de lengte van het middenstuk, het type laatste zinsplaats (complement van V, complement van N, adjunct/geen PVCB), het type pre-eindgroep (voorafgaande groene eindgroep [PART+AUX], voorafgaande rode eindgroep [AUX+PART], geen voorafgaande

eindgroep) en de frequentie van het deelwoord een statistisch significante invloed uit te oefenen op de keuze voor de woordvolgorde.

Opvallend door haar afwezigheid is de enige echte prosodische variabele die in dit proefschrift bestudeerd is, met name de afstand tussen het deelwoordaccent en het laatste preverbale en eerste postverbale accent. Uit die vaststelling lijken we te kunnen concluderen dat in onze dataset met formeel geschreven Belgisch Nederlands geen accentueringsprincipe werkzaam is als onderliggend mechanisme bij de woordvolgordebepaling.¹⁸

Het is uiteraard zo dat er binnen de prosodische dimensie nog andere variabelen zijn waarvan het effect het accentueringsprincipe—zij het indirect—bevestigt: morfologie van het deelwoord, informationaliteit van het linkerwoord, inherentie van het linkerwoord, lengte van het middenstuk, type laatste zinsplaats. De vraag die men zich daarbij moet stellen, is echter in hoeverre het legitiem is om de effecten van die variabelen als een bevestiging van het prosodische principe te beschouwen. Immers, alle variabelen binnen de prosodische dimensie werden voorgesteld als operationalisering van een en hetzelfde principe. Wanneer nu blijkt dat een van die variabelen niet significant tot de verklaring van de geobserveerde waarde bijdraagt, en indien blijkt dat die variabele net de meest specifieke is, dan kan het nog moeilijk volgehouden worden om de andere variabelen in diezelfde prosodische termen te blijven interpreteren. Daarenboven zou men verwachten dat de ‘prosodische’ variabelen die in het globale statistische model wél een effect hebben, minstens gedeeltelijk zouden gelijklopen (i.e. uit het statistische model zouden verdwijnen)—ze testen nl. allemaal een en het zelfde principe!—en dat blijkt niet het geval te zijn. Daardoor dringt de vraag zich op of we de effecten van wat we totnogtoe de indirecte prosodische operationalisering van genoemd hebben niet op een andere manier kunnen interpreteren.

Een alternatieve interpretatie ligt echter niet meteen voor de hand. We hebben dan ook niet de pretentie om in zo’n nieuwe interpretatie te voorzien. In wat volgt, willen we echter wel een mogelijk verklaringskader scheppen waarbinnen een of meerdere van de in dit proefschrift besproken effecten geïnterpreteerd zouden kunnen worden. We zullen de alternatieve verklaring illustreren aan de hand van de effecten van het type grammaticale relatie tussen de laatste zinsplaats en het hoofd

(prosodische dimensie) en de frequentie van het deelwoord (psycholinguïstische dimensie; een effect dat totnogtoe onverklaard bleef). We willen er in ieder geval op voorhand al op wijzen dat het alternatief dat we dadelijk zullen voorstellen (i) in een embryonale fase zit (en daarom nog omstandig geverifieerd zal moeten worden in toekomstig onderzoek) en (ii) in geen geval opgevat mag worden als de ultieme verklaring die andere types verklaringen bij voorbaat uitsluit. Gegeven het multifactoriële karakter van de woordvolgordevariatie gaan we ervan uit dat verschillende mechanismen simultaan werkzaam kunnen zijn en dat daarom ook simultaan verschillende verklaringsprincipes geldig kunnen zijn.

7.3.2 Naar een psycholinguïstische verklaring van de woordvolgordevariatie

7.3.2.1 Uitgangspunt

Bij de zoektocht naar een alternatieve verklaring voor het effect van de deelwoordfrequentie en de effecten van de grammaticale relatie tussen de laatste zinsplaats en het hoofd nemen we de interpretatie van de registereffecten in hoofdstuk 3 als uitgangspunt. Daar stelden we vast dat de modus van het taalgebruik (gesproken versus geschreven), de mate van interactie en de mate van redactionele controle een significant effect hebben op de keuze voor de woordvolgorde:

- Taalgebruikers gebruiken vaker de groene volgorde [PART+AUX] in gesproken taal dan in de geschreven taal.
- Werkwoordelijke eindgroepen die geproduceerd worden in omstandigheden waar twee of meer gesprekspartners met elkaar interageren (gesproken dialogen, geschreven dialogen op babbelkanalen), vertonen significant vaker groene volgordes [PART+AUX] dan in monologische situaties (gesproken monologen, krantenartikelen).
- Werkwoordelijke eindgroepen die geproduceerd zijn in omstandigheden waar de taalgebruiker door de aard van de communicatie niet de mogelijkheid heeft om het resultaat van de taalproductie aan te passen, vertonen vaker groene volgordes [PART+AUX] dan in situaties waar het taalgebruik beperkt gecontroleerd wordt; taaluitingen die ruim gecontroleerd worden, vertonen op hun beurt dan weer minder groene volgordes [PART+AUX] dan beperkt gecontroleerde uitingen. Met andere woorden: hoe groter de mate van

redactionele controle, hoe groter de voorkeur voor de rode woordvolgorde [AUX+PART].

Die drie registreffecten werden geïnterpreteerd in termen van productiedruk: hoe hoger de productiedruk, hoe meer groene woordvolgordes [PART+AUX] er geproduceerd werden, i.e. hoe hoger de temporele restricties op de fonologische en prosodische encoding van de geïntendeerde boodschap en/of op de grammaticale encoding (lexicale selectie en morfologische en syntactische processen), des te meer eindgroepen in de groene woordvolgorde [PART+AUX] verschenen.

Die interpretatie van de resultaten in hoofdstuk 3 zou ons ertoe kunnen leiden om aan te nemen dat de groene woordvolgorde [PART+AUX] de basisvolgorde is, in die zin dat taalgebruikers op de groene volgorde [PART+AUX] terugvallen in situaties waar ze niet de tijd hebben of niet de behoefte voelen om een andere woordvolgorde te kiezen, en dat, omgekeerd, de rode volgorde [AUX+PART] gemarkeerd is (de aandacht wordt naar het hoofdwerkwoord of de volledige eindgroep getrokken) en de groene volgorde [PART+AUX] om sociostilistische redenen vervangt in omstandigheden waar de productiedruk laag of afwezig is.

De assumptie dat de groene volgorde [PART+AUX] de basisvolgorde is, is niet alleen gebaseerd op het registeronderzoek in hoofdstuk 3, ook eerder dialectologisch en taalverwervingsonderzoek wijst in die richting. Zo is in tal van dialectologische onderzoeken (zie hoofdstuk 3) vastgesteld dat (over het algemeen) de groene volgorde [PART+AUX] de dominante volgorde is, waardoor men zou kunnen concluderen dat de variant die het vaakst gebruikt wordt in de meest basale taalvariëteit voor de meest basale communicatiesituaties de basisvariant is. Het taalverwervingsonderzoek van Zuckerman (2001) heeft daarnaast duidelijk gemaakt dat een meerderheid van de kinderen die jonger waren dan 5 jaar in grotere mate de groene volgorde [PART+AUX] kozen dan oudere kinderen en volwassenen. Daaruit zou men kunnen afleiden dat kinderen het eerst de groene volgorde [PART+AUX] verwerven en de groene volgorde [PART+AUX] in die zin dus de basiswoordvolgorde is.¹⁹

7.3.2.2 *Een psycholinguïstische interpretatie van het frequentie-effect*

In paragraaf 6.3 hebben we vastgesteld dat frequente deelwoorden vaker in de rode volgorde [AUX+PART] voorkomen dan minder frequente deelwoorden en dat er in zekere mate een positieve associatie is tussen de frequentie van een deelwoord en de proportie waarin dat deelwoord in de rode volgorde [AUX+PART] voorkomt. Indien men zich daarbij realiseert dat de frequentie waarmee een taaleenheid (een woord of een constructie) voorkomt iets zegt over de mate waarin een taaleenheid cognitief verankerd en toegankelijk is en dus over de snelheid waarmee die taaleenheid verwerkt wordt (identificatie, selectie) in taalproductie- en taalperceptieprocessen—“Higher frequency of a unit or pattern results in a greater degree of what Langacker terms entrenchment, i.e. cognitive routinization, which affects the processing of a unit” (Kemmer & Barlow 2000: x; zie bv. ook Fenk-Oczlon 2001)²⁰—dan kunnen we daaruit besluiten dat het volstrekt in de lijn der verwachting ligt dat frequente deelwoorden vaker in de rode volgorde [AUX+PART] voorkomen:

1. Hoe groter de productiedruk op het encodersproces, i.e. hoe minder tijd er is om de encoding van de geïntendeerde boodschap uit te voeren, hoe meer groene basisvolgordes [PART+AUX].
2. Uit 1 volgt dat de groene basisvolgorde [PART+AUX] sneller/makkelijker te produceren is dan de rode volgorde [AUX+PART]: wat meer basaal is, is beter verankerd en is makkelijker te produceren (linguistic tuning hypothesis: Mitchell et al. 1995).
3. Frequente deelwoorden zijn makkelijker te produceren dan minder frequente deelwoorden (zelfde redeneerschema als in 2).
4. Uit 1, 2 en 3 volgen dat minder frequente deelwoorden vaker in de groene volgorde [PART+AUX] voorkomen en, omgekeerd, frequente deelwoorden vaker in de rode volgorde [AUX+PART]. Bij minder frequente deelwoorden gaat nl. meer tijd naar de selectie van het deelwoord en blijft er aldus minder tijd over om aan stilistische encoding te doen en de neiging om terug te vallen op de makkelijker te produceren groene volgorde [PART+AUX] groter wordt (als een soort compensatiestrategie).

Merk op (i) dat we daarmee niet zeggen dat frequente deelwoorden altijd in de rode volgorde [AUX+PART] voorkomen, maar wel dat er meer tijd is om aan stilistische encoding te doen en daardoor dus de kans groter wordt dat ze in de rode volgorde [AUX+PART] verschijnen; (ii) dat de psycholinguïstische interpretatie van het frequentie-effect van een fundamenteel andere aard is dan het registreffect. In het tweede geval is er nl. sprake van een externe druk op het encoderingsproces (m.n. de beschikbare tijd om een uiting te produceren), terwijl het in het eerste geval om een interne druk gaat, meer bepaald om cognitieve mechanismen die vertraagd of versneld worden door, in dit geval, de verankerings- of toegankelijkheidsgraad van het deelwoord; (iii) dat de intern-temporele restricties op het encoderingsproces (het kost meer tijd om een minder frequent deelwoord te selecteren) gelijkgeschakeld kunnen worden met wat in de psycholinguïstiek bekend staat als *cognitive cost* of *processing cost* (zie bv. Clifton & Duffy 2001); (iv) dat de psycholinguïstische interpretatie enkel volgehouden kan worden, indien aangetoond kan worden dat de geïntendeerde boodschap niet volledig geëncodeerd is vooraleer aangevangen wordt met de articulatie of het schrijven ervan en dat de geheugencapaciteiten niet onbeperkt zijn:

1. De assumptie dat taalgebruikers bij een verhoogde externe of interne druk terugvallen op de basiswoordvolgorde kan pas volgehouden worden als aangetoond kan worden dat taalgebruikers de geïntendeerde boodschap niet volledig grammaticaal encoderen vooraleer aan te vangen met het articulatie- of schrijfproces, i.e. de component die zorgt voor de articulatie of het schrijven opereert slechts op een gedeeltelijk grammaticaal geëncodeerde zin. Anders zou de structuur van de zin al volledig vastliggen vooraleer de articulatie- of schrijfcomponent de output van de grammaticale component begint te verwerken en dan zouden de (intern- of extern-) temporele restricties op het articulatie- of schrijfproces geen invloed kunnen hebben op de grammaticale eigenschappen van de zin. Dat zou bovendien ook impliceren dat moeilijkheden bij de grammaticale productie van een constituent geen invloed kunnen hebben op de grammaticale encoding van een volgende constituent. Experimenten hebben nu aangetoond dat sprekers de geïntendeerde boodschap inderdaad niet volledig grammaticaal encoderen voor ze beginnen te articuleren of schrijven: “Speakers may rarely know exactly how their sentences will end before they begin them [...]” (Bock & Levelt 1994: 971; zie

ook Arnold et al. 2000, Wasow 2002: 46). Hartsuiker & Westenberg 2000 hebben experimenteel kunnen vaststellen dat dat evenzeer geldt voor de geschreven productie als voor de gesproken productie, ondanks de verschillende tijdsvereisten die aan de beide processen ten grondslag liggen.

2. De assumptie dat een taalgebruiker de basiswoordvolgorde kiest in omstandigheden van zware interne of externe productiedruk kan alleen volgehouden worden als aangetoond kan worden dat bepaalde linguïstische structuren meer cognitieve kosten vergen dan andere en dat taalgebruikers geen onbeperkte geheugencapaciteit hebben. Anders zou uiteraard niet verklaard kunnen worden waarom het verwerken van een minder verankerd deelwoord een invloed heeft op de productie van de werkwoordelijke eindgroep. Experimenteel onderzoek naar zgn. *processing overload effects* kunnen als argument dienen om aan te tonen dat taalgebruikers niet beschikken over onbeperkte geheugencapaciteiten (zie bv. de centraal ingebedde relatiefzinnen in Gibson 1998)²¹.

Op basis van de resultaten uit vroeger experimenteel onderzoek kan met andere woorden aannemelijk gemaakt worden dat de verankeringsgraad van een deelwoord wel degelijk een invloed kan hebben op de woordvolgorde in de werkwoordelijke eindgroep.

7.3.2.3 *Een psycholinguïstische interpretatie van het effect van de grammaticale relatie tussen de laatste zinsplaats en het hoofd*

Wanneer we vervolgens de blik wenden naar het effect van de grammaticale relatie van de laatste zinsplaats, dan moeten we vaststellen dat daarover, anders dan voor het frequentie-effect, bij ons weten geen psycholinguïstische informatie beschikbaar is, zodat de verklaring die hierna zal volgen nog meer dan de psycholinguïstische verklaringen hierboven als een hypothese beschouwd moet worden.

We brengen eerst nog in herinnering dat ons corpusgebaseerd onderzoek (paragraaf 4.4.3.2) uitgewezen heeft dat postverbaal gerealiseerde complementen in vergelijking met postverbale adjuncten vaker voorafgegaan worden door een groene volgorde [PART+AUX]. In de lijn van de psycholinguïstische verklaringen die hierboven gegeven zijn voor het register- en frequentie-effect zou men kunnen

aannemen dat de cognitieve kost zwaarder is als complementen, die geselecteerd worden door het hoofd en daardoor belangrijke syntactische, semantische en pragmatische informatie bevatten, langer in het geheugen gehouden moeten worden (m.n. tot de postverbale positie), in vergelijking met perifere adjuncten. Als een relatief groter aandeel van de cognitieve vermogens daaraan gespendeerd wordt, blijft er minder over voor een sociostilistische encodering, zodat de kans groter is dat postverbale complementen voorafgegaan worden door eindgroepen in de makkelijker te produceren groene volgorde [PART+AUX] en adjuncten door eindgroepen in de rode volgorde [AUX+PART]. Een dergelijke interpretatie van het effecten van de grammaticale relatie gaat ervan uit (de assumpties die bij de bespreking van het frequentie-effect aan bod gekomen zijn, worden hier niet meer herhaald):

1. Dat het verschil tussen complementen en adjuncten psychologisch reëel is en dat, met andere woorden, in het mentale lexicon of de mentale grammatica van de taalgebruikers voor elk werkwoord het subcategorisatieschema opgenomen is, i.e. het aantal en het type complementen dat het werkwoord selecteert. Dat een dergelijke assumptie niet zomaar een theoretische constructie projecteert op de/het mentale grammatica/lexicon van de taalgebruikers, blijkt uit het feit dat een heel aantal psycholinguïstische zinsproductiemodellen expliciet gebruik maakt van dergelijke subcategorisatieframes. Die modellen zijn niet zelden gebaseerd op psycho-experimenteel onderzoek naar zgn. *garden path*-zinnen en *late-closure*-ambigüïteiten, waarin duidelijk geworden is dat taalgebruikers gebruik maken van subcategorisatie-informatie om zinnen te parseren (voorbeeld uit Clifton & Duffy 2001: 178):

(56) When Madonna sings the song is a hit.

De NP *the song* wordt initieel geïnterpreteerd als het directe object van het werkwoord *sing*, maar moet gereanalyseerd worden als het subject van de volgende zin bij het parseren van *is*. Die reanalyse, geassocieerd met langere leestijden, toont aan dat taalgebruikers geïnternaliseerde representaties hebben van het type subcategorisatieframes zoals die in de klassieke syntactische theorieën gangbaar zijn en, wat meer is, tijdens het lezen verwachtingen hebben over wat nog moet komen.

2. Dat complementen in het geheugen houden meer van het cognitieve vermogen vergt dan adjuncten. Dat die assumptie niet zomaar een ad-hoc-oplossing is voor de interpretatieve problemen in deze studie, maar een reële veronderstelling, kan aangetoond worden door naar de *Syntactic Prediction Locality Theory* van Gibson (1998) te verwijzen. De SPLT zegt nl. dat “there is a memory cost associated with remembering each category that is required to complete the current input string as a grammatical sentence” (Gibson 1998: 13).

7.3.2.4 Slotbeschouwingen

Aan de hand van de psycholinguïstische interpretatie die in hoofdstuk 3 gegeven werd om het effect van het register te verklaren, hebben we getracht om de effecten van de deelwoordfrequentie en de grammaticale relatie psycholinguïstisch te verklaren. De basisstelling is dat de rode woordvolgorde [AUX+PART] een stilistische optie is en dat de groene woordvolgorde [PART+AUX] gekozen wordt wanneer er zware intern-en/of extern-temporele beperkingen zijn op de taalproductie of wanneer er geen behoefte is om de uiting stilistisch te encoderen. We kunnen de interactie tussen de psycholinguïstische en stilistische motieven bij de keuze voor de rode [AUX+PART] of de groene volgorde [PART+AUX] dan ook als volgt schematisch voorstellen:

Tabel 7.6: Interactie tussen psycholinguïstische en stilistische motieven bij de keuze voor de rode [AUX+PART] of de groene volgorde [PART+AUX].

	Stilistisch: Er is geen behoefte	Stilistisch: er is behoefte
Psycholinguïstisch: er zijn geen beperkingen	Groene volgorde [PART+AUX]	Rode volgorde [AUX+PART]
Psycholinguïstisch: er zijn beperkingen	Groene volgorde [PART+AUX]	Groene volgorde [PART+AUX]

Het spreekt voor zich dat de psycholinguïstische verklaring als een hypothese (of een reeks hypotheses) beschouwd dient te worden en dat veel psycholinguïstisch onderzoek nodig zal zijn om alle aspecten van die interpretatie te verifiëren. In wat voorafging, hebben we al een heel aantal assumpties die aan de interpretatie ten

grondslag liggen aan de hand van eerdere resultaten van psycho-experimenteel onderzoek aannemelijk trachten te maken. Tot de zaken die in toekomstig onderzoek zeker nog behandeld zullen moeten worden, behoren de volgende vier:

1. Waarom zou een taalgebruiker, die geen gebruik kan maken van onbeperkte cognitieve capaciteiten, ervoor opteren om een complement dat evengoed in het middenstuk kan voorkomen te extraponeren als de cognitieve kost veel hoger ligt (het moet nl. langer in het geheugen gehouden worden)? Het is duidelijk dat een gedeeltelijk antwoord op die vraag gevonden kan worden in het vermijden van een al te complex middenstuk (Arnold et al. 2000, Wasow 2002), maar dan nog moet de vraag gesteld worden hoe die beide tegengestelde krachten—een kracht die ernaar streeft om zo weinig mogelijk te extraponeren en een kracht die ernaar streeft om het middenstuk niet te complex te maken—zich in taalproductieomstandigheden ten opzichte van elkaar verhouden.
2. Zelfs al zou het correct zijn om te veronderstellen dat postverbale complementen meer cognitieve kosten vereisen en dus minder snel geproduceerd kunnen worden dan postverbale adjuncten en dat de geheugencapaciteit niet onbeperkt is, dan nog volgt daar niet onmiddellijk uit dat er geen geheugencapaciteit meer overblijft om een rode woordvolgorde [AUX+PART] te produceren.
3. Uit de bovenstaande argumentatie is geen onderscheid gemaakt tussen facultatieve complementen en obligatoire complementen. Aangezien het eerste type niet altijd gerealiseerd hoeft te worden in een concrete uiting (en dus ook niet altijd in het geheugen gehouden hoeft te worden) en de tweede wel, kan men zich de vraag stellen of het verschil tussen die beide types een verschil in productiekost met zich meebrengt.
4. Totnogtoe hebben we de psycholinguïstische hypothese enkel in termen van productiekosten geformuleerd. Men kan zich uiteraard ook de vraag stellen of de argumentatie hierboven geprojecteerd kan worden op perceptieprocessen, waarbij men meer concreet zou aannemen dat de groene volgorde [PART+AUX] de perceptie (i.e. de leessnelheid) bevordert.²²

In toekomstig onderzoek zal ook nagedacht moeten worden of het mogelijk is om de andere effecten die in dit proefschrift aan bod gekomen zijn op dezelfde psycholinguïstische manier te interpreteren.

Noten:

1. In de vorige hoofdstukken zijn de redundantie- en interactieverbanden in zeker mate al bestudeerd voor sommige factoren binnen een bepaalde dimensie, maar het moge duidelijk zijn dat dergelijke verbanden ook tussen factoren over dimensiegrenzen heen kunnen plaatsvinden.
2. Voor de n-gecontroleerde variabele geldt dat eindgroepen die niet voorafgegaan worden door een nomen de LLR-waarde 0 krijgen (= geen vaste verbinding).
3. Zowel de n-gecontroleerde als de niet-n-gecontroleerde variabele zijn opgenomen. Voor de niet-n-gecontroleerde variabele geldt hetzelfde als wat in noot 2 gezegd werd.
4. Voor de eindgroepen met een deelwoord en het hulpwerkwoord *hebben* en *worden*, die geen adjectiviteitswaarde toegekend kregen, omdat voorlopig het traditionele standpunt ingenomen is dat deelwoorden met *hebben* of *worden* enkel verbaal geïnterpreteerd kunnen worden, werd de adjectiviteitswaarde 0 (=extreem verbaal) toegekend.
5. Hiervoor is de uitgebreide operationalisatie gebruikt, i.e. twee- en meerledige pre-eindgroepen, samengevat in twee basistypes.
6. In plaats van de wat informelere term *factor* zullen we vanaf hier systematisch spreken van *verklarende variabele* of *predictorvariabele*.
7. Dat wil zeggen dat de CBA geen bepaalde onderliggende verdeling van de data veronderstelt.
8. Omdat een CBA telkens een binaire splitsing wil maken, zoekt ze in het geval van een continue variabele het meest ideale snijpunt. In het geval van een categorische variabele met meer dan twee waarden wordt in een hergroepering voorzien.
9. Deviantie is een statistische maat die—toegepast op ons geval hier—aangeeft in welke mate een welbepaalde splitsing de geobserveerde gegevens benadert. Hoe beter een splitsing de gegevens benadert, hoe kleiner de deviantie, zodat de splitsing met de kleinste deviantie (dus: de splitsing met de grootste reductie in deviantie) de meest optimale splitsing is (zie Harrell (2001: 37) voor alternatieven).
10. Voor een bevattelijke conceptuele inleiding tot de logistische regressie, toegepast op een linguïstische case study, verwijzen we naar Grondelaers et al. 2001.
11. Zoals de naam zelf het zegt, zijn GLM's een uitbreiding van het algemene lineaire model. Die uitbreiding zorgt ervoor dat niet alleen responsvariabelen met een normale verdeling, maar ook bv. binomiaal verdeelde responsvariabelen, zoals de woordvolgorde in tweeledige werkwoordelijke eindgroepen, in aanmerking komen om statistisch gemodelleerd te worden. Voor een overzicht van andere GLM's, zie Agresti (1996: 71-102) en Paolillo (2002: 175-189).

12. De interpretatie van de regressiecoëfficiënten kan dan in termen van conditionele o.r.'s gebeuren. De regressiecoëfficiënten kunnen makkelijk omgezet worden naar o.r.'s (zie hierna), waardoor het verschil tussen een parameterwaarde en de referentiewaarde geïnterpreteerd kan worden als een conditionele o.r., aangezien het effect van de andere parameters constant gehouden wordt (zoals in de CMH-analyses). We komen hier dadelijk op terug.
13. In dit verband merken we nog op dat we de drie variabelen *aan- vs. afwezigheid van een laatste zinsplaats*, *type hoofd van de laatste zinsplaats* en *grammaticale relatie tussen het hoofd en de laatste zinsplaats* om technische redenen (dataschaarste) in één variabele ondergebracht hebben. De waarden die we onderscheiden, zijn: complement van het hoofdwerkwoord, complement van een preverbale constituent of adjunct/geen PVCB.
14. De c-maat is equivalent aan de *Receiver Operating Characteristic* (zie Harrell 2001: 247).
15. Omdat het logistische-regressiemodel gebaseerd is en getest wordt op een en dezelfde dataset, is het niet ondenkbaar dat de c-waarde hoger ligt dan het geval zou zijn wanneer het model op een onafhankelijke testsample getest zou worden. Bootstrapvalidatie is een techniek die het gevaar van een dergelijk overoptimisme reduceert: uit de dataset waarop het model gefit wordt, wordt 100 keer een sample getrokken en voor elke sample wordt de c-waarde berekend. Vervolgens wordt de gemiddelde c-waarde over die 100 samples berekend (Harrell 2001: 87-90, 94-97).
16. Een vaak gebruikte maat voor predictieve kracht is het aantal correct geclassificeerde items. De datapunten worden daarbij op basis van de voorspelde probabiliteit dat $Y=1$ (een cijfer tussen 0 en 1) in twee groepen gedeeld, een groep met $Y=1$ (i.c. de rode volgorde [AUX+PART]) en een groep met $Y=0$ (de groene volgorde [PART+AUX]). Het cut-off-punt kan daarbij zelf gekozen worden (bv. 0.5). Ondanks het feit dat de resultaten makkelijker interpreteerbaar zijn (percent correct voorspelde responsen), volgen we Harrell (2001: 248-249) in zijn afwijzing van de maat.
17. In dat verband is het niet onbelangrijk om op te merken dat in het kader van dit doctoraatsproject ook nog een aantal andere variabelen onderzocht is, m.n.: lengte van het deelwoord in aantal syllabes; reflexiviteit; negatie; type, functie en lengte van de laatste preverbale constituent; frequentie van het laatste preverbale woord; syllabekwaliteit van het laatste preverbale woord en deelwoord. Aangezien een voorbereidende logistische-regressieanalyse uitwees dat die variabelen geen bijdrage leveren aan de verklaring van de geobserveerde variatie, hebben we ervan afgezien om er in dit proefschrift over te rapporteren.
18. (i) De conclusie laat de mogelijkheid open dat een accentueringsprincipe, zoals dat door De Schutter (1996) geformuleerd is, bijvoorbeeld in gesproken taal wel een rol kan spelen (zie bv. het onderzoek van Swerts 1998). (ii) Daarnaast moet ook opgemerkt worden dat de operationalisering van de accentueringshypothese niet perfect strookt met de initiële ideeën van De Schutter, waardoor de mogelijkheid opengehouden moet worden dat een specifiekere operationalisering wel een effect van de accentdistributie vindt. (iii) Dat het accentueringsprincipe niet werkzaam zou zijn in het formele geschreven Belgische Nederlands sluit uiteraard niet uit dat een ander prosodisch principe de keuze van de woordvolgorde kan beïnvloeden.

19. We gaan er ondanks de methodologische bezwaren die in hoofdstuk 1 geuit zijn (het beperkte aantal experimentele items, onduidelijke selectie van de testpersonen en het soms onorthodoxe gebruik van statistische technieken) vanuit dat de grote tendensen uit Zuckermans onderzoek correct zijn.
20. Dat zoiets niet alleen een theoretische aanname is, maar daadwerkelijk de taalproductie en -perceptie beïnvloedt, is al meermaals aangetoond in psycholinguïstisch onderzoek (zie bv. Mitchell et al. 1995).
21. Merk op dat Gibson 1998 en Clifton & Duffy 2001 zich enkel concentreren op zinsperceptie. De algemene stellingen hier kunnen echter makkelijk toegepast worden op taalproductieprocessen (zie bv. Arnold et al. 2000, Wasow 2002).
22. Uit een eerste psycho-experimentele test blijkt alvast dat de perceptie inderdaad in sommige omstandigheden beïnvloed wordt door de keuze voor de groene [PART+AUX] of de rode volgorde [AUX+PART]. Wegens het onvoldragen karakter van het experiment, zien we er echter van af om de volledige psycho-experimentele resultaten hier te presenteren.

Hoofdstuk 8

Conclusies

In dit proefschrift hebben we een van de meest intrigerende fenomenen van syntactische variatie bestudeerd die het Nederlands rijk is: de werkwoordsvariatie in de tweeledige werkwoordelijke eindgroep met een deelwoord en een hulpwerkwoord. Meer bepaald hebben we een antwoord gezocht op de vraag *wat* de keuze voor de groene [PART+AUX] of de rode [AUX+PART] volgorde in het concrete taalgebruik bepaalt en *waarom* we überhaupt over twee volgordealternatieven beschikken. We hebben daartoe aan de hand van een uitgebreide dataset met reëel geattesteerd taalmateriaal uit twee goed gecontroleerde en representatieve corpora van het Nederlands de invloed van een hele set taalinterne en taalexterne factoren geanalyseerd. Om de studie daarnaar georganiseerd te laten verlopen, hebben we de verschillende factoren, gebaseerd op wat we wisten uit vroeger onderzoek, ingedeeld in vier dimensies: een contextuele dimensie, een prosodische dimensie, een semantische dimensie en een psycholinguïstische dimensie.

Langs de contextuele dimensie hebben we de invloed van de regio en het taalregister bestudeerd (hoofdstuk 3), langs de prosodische dimensie hebben we onderzocht of de keuze tussen de groene [PART+AUX] en de rode volgorde [AUX+PART] afhankelijk is van de accentdistributie in en rond de werkwoordelijke eindgroep (hoofdstuk 4), langs de semantische dimensie hebben we de volgordevariatie in verband gebracht met het onderscheid tussen adjectivische en (verschillende types) werkwoordelijke deelwoorden (hoofdstuk 5) en langs de psycholinguïstische dimensie werd tot slot de rol van taalproductieprocessen onder de loep genomen (hoofdstuk 6). We vatten voor de verschillende dimensies de opmerkelijkste resultaten puntsgewijs samen:

**Contextuele
dimensie**

- In het Belgische Nederlands wordt vaker de groene volgorde [PART+AUX] gebruikt dan in het Nederlandse Nederlands.
- In de Brabantse regio wordt vaker de rode volgorde [AUX+PART] gebruikt dan in de rest van Vlaanderen; in Oost-Vlaanderen wordt vaker de groene volgorde [PART+AUX] gebruikt dan in de andere Vlaamse provincies.
- In gesproken taalgebruik wordt vaker de groene volgorde [PART+AUX] gebruikt dan in geschreven taalgebruik.
- In communicatiesituaties met een hoge mate van interactie wordt vaker de groene volgorde [PART+AUX] gebruikt dan in situaties met een beperktere interactie.
- In communicatiesituaties met een hoge mate van redactionele controle wordt vaker de rode volgorde [AUX+PART] gebruikt dan in situaties met een beperktere vorm van redactionele controle.

Prosodische dimensie

- Scheidbaar samengestelde deelwoorden komen vaker in rode eindgroepen [AUX+PART] voor dan niet-scheidbaar samengestelde deelwoorden.
- Voorafgaande indefiniete substantiefgroepen doen, in vergelijking met definiete substantiefgroepen en andere woordsoorten, de voorkeur voor de rode volgorde [AUX+PART] stijgen.
- Voorafgaande hoog-informationele woorden doen, in vergelijking met laag-informationele woorden, de voorkeur voor de rode volgorde [AUX+PART] stijgen.
- Een voorafgaand niet-werkwoordelijk deel van een vaste verbinding doet, in vergelijking met de afwezigheid van een dergelijk niet-werkwoordelijk deel van een vaste verbinding, de voorkeur voor de rode volgorde [AUX+PART] stijgen.

Prosodische dimensie

- Een lang middenstuk doet, in vergelijking met een kort middenstuk, de voorkeur voor de rode volgorde [AUX+PART] stijgen.
- De aanwezigheid van een laatste zinsplaats doet, in vergelijking met zinnen zonder laatste zinsplaats, de voorkeur voor de groene volgorde [PART+AUX] stijgen.
- Postverbale complementen doen, in vergelijking met zinnen met een postverbaal adjunct, de voorkeur voor de groene volgorde [PART+AUX] stijgen.
- Een laatste zinsplaats die afhankelijk is van het deelwoord doet, in vergelijking met een laatste zinsplaats die afhankelijk is van een preverbale constituent, de voorkeur voor de groene volgorde [PART+AUX] stijgen.
- Naarmate er minder onbeklemtoonde syllabes tussen het deelwoordaccent en het laatste preverbale accent staan, stijgt de voorkeur voor de rode volgorde [AUX+PART].
- Naarmate er minder onbeklemtoonde syllabes tussen het deelwoordaccent en het eerste postverbale accent staan, stijgt de voorkeur voor de groene volgorde [PART+AUX].

Semantische dimensie

- Adjectivische(re) deelwoorden komen vaker in de groene volgorde [PART+AUX] voor dan werkwoordelijke(re) deelwoorden.
- Het is niet uitgesloten dat zuiver adjectivische deelwoorden in de rode volgorde [AUX+PART] verschijnen.
- Een toestandsgericht deelwoord is niet incompatibel met de rode volgorde [AUX+PART], meer nog, het kan zelfs typisch zijn voor de rode volgorde [AUX+PART].
- Eindgroepen met het hulpwerkwoord *zijn* komen vaker dan respectievelijk *worden* en *hebben* in de groene volgorde [PART+AUX] voor.
- Passieve constructies komen vaker in de groene volgorde [PART+AUX] voor dan actieve constructies.

- De volgorde van de werkwoorden in een voorafgaande werkwoordelijke eindgroep bepaalt de volgorde van de eindgroep die erop volgt: indien een eindgroep in de rode volgorde [AUX+PART] voorafgaat, dan wordt die volgorde overgenomen in de volgende eindgroep; indien een eindgroep in de groene volgorde [PART+AUX] voorafgaat, dan wordt die volgorde overgenomen in de volgende eindgroep.
- De mate van beïnvloeding van de voorafgaande werkwoordelijke eindgroep is niet afhankelijk van de afstand tussen de twee eindgroepen.
- Het is niet uitgesloten dat in puzzelexperimenten een onbedoeld syntactisch primingeffect aanwezig is.
- Deelwoorden met een hoge frequentie komen vaker in de rode volgorde [AUX+PART] voor dan deelwoorden die minder frequent zijn.

Ook al lijkt de voorliggende studie op het eerste gezicht in weinig te verschillen van de vele (empirische) studies die in het verleden al verschenen zijn, toch menen we dat we met dit proefschrift een belangrijke stap voorwaarts gezet hebben in de zoektocht naar het hoe en waarom van de woordvolgordevariatie. Dat komt niet zozeer doordat we een gigantische inhoudelijke doorbraak geforceerd zouden hebben, maar wel doordat we door de methodologische en analytische gestrengheid waarmee we te werk gegaan zijn in zekere zin vaste grond onder de voeten gekregen hebben. We hebben er nl. voor gezorgd dat de werkwoordelijke eindgroepen aan gedetailleerde selectiecriteria voldoen; dat linguïstische hypothesen op een transparante en bruikbare manier geoperationaliseerd zijn; dat de beschikbare gegevens volledig en systematisch geanalyseerd zijn; dat de invloed van de verschillende factoren op een en dezelfde dataset geanalyseerd zijn; dat de resultaten van die analyses met behulp van een batterij aan statistische tests op hun betrouwbaarheid en representativiteit gecontroleerd zijn.

Dat die methodologische en analytische eigenschappen van de voorliggende studie niet zomaar epifenomenen zijn van empirisch taalwetenschappelijk onderzoek, is eerder al gebleken uit de bespreking van de resultaten van vroeger wetenschappelijk onderzoek (paragraaf 1.3.4): zonder een volledige duidelijkheid op het vlak van de

methodologische en analytische fundamente van het onderzoek, is een accurate interpretatie van de resultaten onmogelijk; zonder een volledige duidelijkheid op het vlak van de methodologische en analytische fundamente van het onderzoek wordt de mogelijkheid van replicatie de pas afgesneden. Aangezien wij meer toeneiging hebben voor een linguïstiek van coöperatie en interactie dan voor een linguïstiek van cocooning, wil dit proefschrift een krachtig pleidooi zijn voor wat Geeraerts (2004) een tragere, strengere en cumulatievere linguïstiek noemt.

Het is duidelijk dat een dergelijke manier van werken een niet geringe tijdsinvestering vergt, zoals in de loop van dit proefschrift gebleken mag zijn (datagaring, datapurificatie, data-annotatie, data-analyse). De winst die ze echter oplevert, is o.i. vele malen groter. Zo zijn we er in geslaagd om notoir moeilijk vatbare linguïstische categorieën als *categoriale status van het deelwoord* en *vaste verbinding* te operationaliseren, zodat ze in een empirische studie systematisch en objectief gebruikt kunnen worden. We hebben het effect van de verschillende factoren geanalyseerd op een en dezelfde dataset, wat het voordeel heeft dat de kracht en de richting van die factoren rechtstreeks met elkaar vergeleken kunnen worden. Daarenboven zijn de lokale relaties tussen de verschillende factoren in extenso onderzocht (bv. Cochran-Mantel-Haenszel-analyses) en zijn we er voor het eerst in geslaagd een geïntegreerde statistische verklaring te bieden voor de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep (logistische-regressie- en classificatieboomanalyse). De resultaten van die geïntegreerde statistische verklaring hebben ons ertoe gebracht de woordvolgordevariatie op een andere manier te verklaren dan totnogtoe gebeurd is, met name in termen van mate van productiedruk. Zonder daarbij te willen beweren dat productiedruk de enige verklaring is om de alternatie tussen de beide volgordes te verklaren, is daarmee een nieuwe en o.i. plausibele verklaringshypothese geformuleerd die door haar eenvoud en heldere formulering uitnodigt tot verificatie.

Het spreekt voor zich dat de zoektocht naar de woordvolgordevariatie met deze studie niet afgelopen is. Wel integendeel! In de loop van dit proefschrift moge gebleken zijn dat heel wat pistes voor verder onderzoek openliggen. We sommen er een aantal op. In de eerste plaats zal toekomstig onderzoek de resultaten van onze studie moeten verifiëren, en eventueel verder verfijnen en aanvullen. Er zal in de

tweede plaats nagedacht moeten worden of het huidige complex van factoren gecompleteerd kan worden door een aantal andere variabelen die hier niet bestudeerd zijn. In de derde plaats zal de globale statistische analyse zeker overgedaan moeten worden voor een model waarin ook taalexterne factoren geïncorporeerd zijn en er zal nagegaan moeten worden in hoeverre dat globale statistische model gemodificeerd moet worden (rekening houdend met bijvoorbeeld de interacties tussen de verschillende factoren). Andere (multivariate) statistische technieken kunnen daarbij een welkome aanvulling bieden op de technieken die in dit proefschrift gebedigd zijn. Tot slot zal in vervolgonderzoek de psycholinguïstische interpretatie die we geboden hebben psycho-experimenteel geverifieerd moeten worden.

Appendices

Appendix A: lijsten van geconsulteerde algemene literatuur

Appendix A biedt een thematisch overzicht van een aantal niet-wetenschappelijke, algemene werken die geraadpleegd werden om een idee te krijgen in hoeverre het fenomeen van de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep met een deelwoord en een hulpwerkwoord in de taalkundige wereld en de taalgemeenschap verankerd is (zie paragraaf 1.3). Achtereenvolgens komen grammatica's en inleidingen in de Nederlandse syntaxis (A.1), grammatica's en inleidingen in het Nederlands voor anderstaligen (A.2), woordenboeken van grammaticale termen op het gebied van de Nederlandse taalkunde (A.3) en stijlboeken, taaladvies-, taalverzorgings- en taalzuiveringswerken (A.4) aan bod. Een asterisk voor de referentie betekent dat het werk in kwestie de woordvolgordevariatie in de tweeledige werkwoordelijke eindgroep vermeldt.

A.1 Grammatica's en inleidingen in de Nederlandse syntaxis

**ANS – Algemene Nederlandse Spraakkunst* (1984). G. Geerts, W. Haeseryn, J. de Rooij, M.C. Van den Toorn (red.). Groningen / Leuven: Wolters-Noordhoff / Wolters.

**ANS – Algemene Nederlandse Spraakkunst* (1997). W. Haeseryn, K. Romijn, G. Geerts, J. de Rooij, M.C. Van den Toorn (red.). Groningen / Deurne: Martinus Nijhoff / Wolters Plantyn.

Balk-Smit Duyzentkunst, F. (1994). *De woorden en hun zin. Grammatica voor iedereen*. Amsterdam: Nijgh & Van Ditmar.

- Balk-Smit Duyzentkunst, F. (2000). *Grammatica van het Nederlands*. Den Haag: Sdu.
- Bart, P. Van, J. Kerstens & A. Sturm (1998). *Grammatica van het Nederlands. Een inleiding*. Amsterdam: Amsterdam University Press.
- Beckers, P., G. de Vletter & P. Vloet (1979). *Eenvoudige algemene grammatica*. Zutphen: Thieme.
- (*Berg, B. Van den (1957). *Beknopte Nederlandse Spraakkunst*. 's Gravenhage: van Goor.
- *Berg, B. Van den (1967). *Beknopte Nederlandse Spraakkunst*. Den Haag/Brussel: van Goor.
- Berg, B. Van den (1978). *De grammatica van het Standaardnederlands*. Groningen: Wolters-Noordhoff.
- Calcar, W.J.M. Van (1983). *Een nieuwe grammatica voor taalbeschouwing en taalbeheersing*. Leuven: Acco.
- *Cosijn, P.J. & J. Te Winkel (1888). *Nederlandsche spraakkunst: syntaxis*. Haarlem: Bohn.
- *Daenen, C. & H. De Jonghe (1978). *Nederlandse spraakkunst*. Lier: Van In.
- Droste, F.G. (1965). *Grondbeginselen van de Nederlandse grammatica*. Den Haag: Van Goor.
- Es, G.A. Van (1966). *Nederlandse syntaxis in klein bestek*. Zwolle: Tjeenk Willink.
- Fokkema, G. (1981). *Mini-cursus: beknopte handleiding voor de ontleding met bijlagen*. Amsterdam: VU
- *Fraussen, E. (1924). *Uitgebreide Nederlandsche spraakkunst voor de hogere klassen van middelbaar en normaal onderwijs*. Wetteren: De Meester.
- *Fraussen, E. (1935). *Beknopte Nederlandsche spraakkunst voor middelbaar onderwijs*. Mechelen: St. Franciscus.
- Groot, A.W. de (1949). *Structurele syntaxis*. Den Haag: Servire.
- *Hasselbach, W.H. (1905). *Nederlandsche-spraakkunst*. Groningen: Wolters.
- *Hertog, C.H. den (1973). *Nederlandse spraakkunst*. Amsterdam: Versluys.
- *Holtvast, K. (1908). *Beknopte Nederlandsche Spraakkunst*. Groningen: Noordhoff.
- *Joos, A.M. (1901). *Vlaamsche spraakleer ten gebruike van het middelbaar en normaal onderwijs*. Gent: Siffer.

- Klooster, W.G., H.J. Verkuyl & J.H.J. Luif (1971). *Inleiding tot de syntaxis: praktische zinsleer van het Nederlands*. Culemborg: Stam-Robijns.
- *Klooster, W. (2001). *Grammatica van het hedendaags Nederlands: een volledig overzicht*. Den Haag: Sdu.
- Kraak, A. & W.G. Klooster (1972). *Syntaxis*. Culemborg: Stam-Robijns.
- Kummer, J.C. (1914). *Nederlandsche spraakkunst*. Amsterdam: Van Looy.
- Langendonck, J.J.M. Van (1861). *Nederlandsche spraakkunst met oefeningen: voorafgegaan van eenige bedenkingen over eigenbelang, godsdienst en nationaliteit, ten aanzien der taal, en gevolgd van eene beknopte woordafleidkunde*. Lier: Van In.
- Ligtenberg, P.H., H. Van Rossum & B.W. Schippers (s.d.). *Nederlands voor de kweekschool*. Groningen: Noordhoff.
- *Lubbe, H.F.A. Van der (1965). *Woordvolgorde in het Nederlands. Een synchrone structurele beschouwing*. Assen: Van Gorcum.
- *Luif, J. (1994). *In verband met de zin: inleiding in de Nederlandse spraakkunst*. Amsterdam: AUP.
- *Luif, J. (1998). *In verband met de zin: inleiding in de Nederlandse spraakkunst*. Amsterdam: AUP.
- *Olbrechts, G. (1966). *De essentie van het Nederlands*. Antwerpen: De Sikkel.
- Overdiep, G.S. (1928). *Moderne Nederlandsche grammatica voor gymnasia, HBS en kweekscholen alsmede voor zelfstandige studie*. Zwolle: Tjeenk Willink.
- *Overdiep, G.S. (1937). *Stilistische grammatica van het moderne Nederlandsch*. Zwolle: Tjeenk Willink.
- *Overdiep, G.S. (1949). *Stilistische grammatica van het moderne Nederlandsch*. Zwolle: Tjeenk Willink.
- *Paardekooper, P.C. (1960). *ABN-spraakunst: voorstudies*. Den Bosch: Malmberg.
- *Paardekooper, P.C. (1963). *Beknopte ABN-syntaxis*. Den Bosch: Malmberg.
- *Paardekooper, P.C. (1966). *Kleine ABN-syntaxis in vraag en antwoord voor VHMO en ander voortgezet onderwijs*. Den Bosch: Malmberg.
- *Paardekooper, P.C. (1971). *Beknopte ABN-syntaxis*. 's Hertogenbosch: Malmberg.

- *Paardekooper, P.C., C. Moeyaert & J. Peperstraete (1975). *Beknopte ABN-spraakunst*. Antwerpen: Scriptoria.
- *Paardekooper, P.C. (1977). *Beknopte ABN-syntaxis*. Eindhoven: Eigen beheer.
- Paardekooper, P.C. (1979). *Kleine ABN-syntaxis in vraag en antwoord voor HAVO, VWO en onderwijzersopleiding*. 's Hertogenbosch: Malmberg.
- Rijpma, E. & F.G. Schuringa (1917). *Nederlandsche spraakkunst ten dienste van kweekscholen, normaallessen en studeerenden voor de hoofdacte*. Groningen: Wolters.
- Rijpma, E. & F.G. Schuringa (1925). *Taalverschijnselen. Aanteekeningen en oefeningen bij de Nederlandsche spraakkunst*. Groningen: Wolters.
- *Rijpma, E., F.G. Schuringa (1957). *Nederlandse spraakkunst*. Groningen: Wolters [Bewerkt door J. Naarding].
- *Rijpma, E., F.G. Schuringa (1978). *Nederlandse spraakkunst*. Groningen: Wolters-Noordhoff [Bewerkt door J. Van Bakel].
- *Schutter, G. De & P. Van Hauwermeiren (1983). *De structuur van het Nederlands: taalbeschouwelijke grammatica*. Malle: De Sikkel.
- *Smedts, W. & W. Van Belle (1993). *Taalboek Nederlands*. Kapellen: DNB Pelckmans.
- *Smedts, W. & W. Van Belle (1997). *Taalboek Nederlands*. Kapellen: Pelckmans.
- *Smedts, W. & W. Van Belle (2003). *Taalboek Nederlands*. Kapellen: Pelckmans.
- Terwey, T. & M. Mieras (1903). *Nederlandsche Spraakkunst*. Groningen: Wolters.
- Tinbergen, D.C. (1970). *Nederlandse spraakkunst*. Culemborg: Tjeenk Willink.
- Toorn, M.C. Van den (1978). *Nederlandse grammatica*. Groningen: Tjeenk Willink.
- Toorn, M.C. Van den (1982). *Nederlandse grammatica*. Groningen: Wolters-Noordhoff.
- *Vindevogel, T. (1968). *Nederlandse grammatica*. Kapellen: De Sikkel.
- *Vooyoys, C.G.N. de & M. Schönfeld (1949). *Nederlandse spraakkunst*. Groningen: Wolters.
- *Vooyoys, C.G.N. de & M. Schönfeld (1967). *Nederlandse spraakkunst*. Groningen: Wolters.

A.2 *Grammatica's en inleidingen in het Nederlands voor anderstaligen*

- Bertha, A. (1973). *Synthèse de grammaire néerlandaise*. Bruxelles: De Boeck.
- Bind, R.B. & W.Z. Shetter (1971). *Een goed begin... A contemporary Dutch reader*. The Hague: Martinus Nijhoff.
- *Brachin, P. (1977). *La langue néerlandaise. Essai de présentation*. Bruxelles: Didier.
- *Brachin, P. & P. Vincent (1985). *The Dutch language. A survey..* Leiden: Brill.
- Donaldson, B.C. (1981). *Dutch reference grammar*. 's Gravenhage: Nijhoff.
- *Fehringer, C. (1999). *A reference grammar of Dutch. With exercises and key*. Cambridge: CUP.
- *Florijn, A.F., J.A. Lalleman & J.H. Maureau (1994). *De regels van het Nederlands: grammatica voor anderstaligen*. Groningen: Wolters-Noordhoff.
- Godin, P., P. Ostyn & F. Degreef (1993). *La pratique du néerlandais avec ou sans maître*. Louvain-la-Neuve : Peeters.
- Kampen, H. Van & R. Stumpel (1995). *Dutch for self-study. Nederlands voor anderstaligen*. Utrecht: Spectrum.
- Lambregtse, C. (1953). *Fundamentals of practical Dutch grammar*. Grand Rapids: Eerdmans.
- Lindenburg, T.W.H. (1960). *Kleine niederländische Sprachlehre für Schul-, Privat- und Selbstunterricht: Methode Gaspey-Otto-Sauer*. Heidelberg: Groos.
- *Meer, M.J. Van der (1923). *Grammatik der neuniederländischen Gemeinsprache (Het algemeen beschaafd)*. Heidelberg: Winter.
- Shetter, W.Z. (1981). *Introduction to Dutch: a practical grammar*. The Hague: Nijhoff.
- Shetter, W.Z. (1994). *Dutch. An essential grammar*. London: Routledge.
- Smit, J. & R.P. Meijer (1958). *Dutch grammar and reader with exercises*. Melbourne: Melbourne University Press.
- *Vandeputte, O, P. Vincent & T. Hermans (1986). *Dutch. The language of twenty million Dutch and Flemish people*. Rekkem: Ons Erfdeel.

A.3 Woordenboeken van grammaticale termen op het gebied van de Nederlandse taalkunde

- *Pensaert, L. & L. Van der Sande (1991). *Grammaticaal en taalkundig lexicon: op basis van de Algemene Nederlandse Spraakkunst*. Antwerpen: Standaard.
- Pollmann, T. & A. Sturm (1978). *Over zinnen gesproken: termen en begrippen van de traditionele grammatica*. Culemborg: Educaboek.
- Vrisekoop, M.C. (2001). *Grammaticale termen: het gereedschap voor elke taal*. Bussum: Coutinho.

A.4 Stijlboeken, taaladvies, taalverzorging, taalzuivering

- *Berg, J. de (2001). *Trouw schrijfboek*. Amsterdam: Muntinga.
- Boeck, A.E. (1922). *Nederlandsch taaleigen: woorden en wendingen of proeve van taalzuivering*. Lier: Van In.
- Boeck, A.E. & J. Grauls (1944). *Nederlandsch taaleigen: woorden en wendingen of proeve van taalzuivering*. Lier: Van In.
- *Boer, H. De (1972). *Schriftelijk rapporteren*. Utrecht: Spectrum.
- Bruycker, G. De, R. Haeseryn & V.F. Vanacker (1995). *Correct Nederlands thuis, in winkel en bedrijf*. Heule: UGA.
- Burger, P & J. de Jong (1997). *Handboek stijl: adviezen voor aantrekkelijk schrijven*. Groningen: Nijhoff.
- *Buyse, P. & W. Penninckx (1977). *Verzorgd taalgebruik*. Heule: UGA.
- *Buyse, P. & W. Penninckx (1984). *Verzorgd taalgebruik*. Heule: UGA.
- *Charivarius (1945). *Is dat goed Nederlandsch?* Antwerpen: Kompas.
- *Charivarius (s.d.). *Is dat goed Nederlandsch?* Antwerpen: Kompas.
- Claes, D. (1891). *Gemengde taal- en letterkundige aanmerkingen*. Gent: Siffer.
- Cockx, P. (1989). *Taalwijzer*. Leuven: Davidsfonds.
- Cockx, P. (2000). *Taalwijzer*. Leuven: Davidsfonds.
- Crombrugge, E. Van, L. Dooms & M.J.L. Van Nierop (1973). *Keurig Nederlands in de administratie... en elders*. Gent: Daphne.

- Daman, J.A. (1946). *Het algemeen beschaafd in Vlaanderen (Zuid-Nederland)*. Gent: Daphne.
- Dezaire, P. (1964). *Ons nederlands: een taal- en stijlboek*. Utrecht: Spectrum.
- *D'Haenens, L, W. Penninckx, P. Van Hauwermeiren, R. Dekegel, L. Dalle & H. Buermans (1991). *In duidelijk Nederlands*. Brussel: Ministerie van de Vlaamse gemeenschap. Bestuur afstandsonderwijs.
- Diemer, W. (1964). *Waarop letten bij het schrijven in het Nederlandsch?* Delden: Stabo.
- Dominicus, F.C. (1962). *Schrijft u ook zulk Nederlands?* Den Haag: Van Goor.
- Douma, P. & R. Geel (1990). *Schrijven voor een groot publiek: handleiding voor het schrijven van teksten voor de media*. Leiden: Nijhoff.
- Edens, K. (1995). *Geachte lezer: lezergerichte zakenbrieven*. Houten: Bohn Stafleu Van Loghum.
- *Eijk, I. Van (1984). *Taalhulp*. Amsterdam: Bakker.
- *Eijk, I. Van (2000). *De beste taaltips: een selectie uit het handboek Taaltips*. Alphen a.d. Rijn: Samsom.
- Elling, M.G.M. (1984). *De techniek van het schriftelijk rapporteren*. Groningen: Wolters-Noordhoff.
- *Gessel, H. Van & J.K. Hulsbosch (1992). *De Volkskrant: Stijlboek*. 's Gravenhage: Sdu.
- Gillaerts, P. (1987). *Handboek normatieve taalbeheersing: gids voor correct Nederlands*. Leuven: Acco.
- Goed woordgebruik in bedrijf en techniek: woordenlijst met taalkundige aanwijzingen* (1994). Delft: Nederlands Normalisatie-Instituut.
- Grauls, J. P. Melis & C.H. Peeters (1933). *Taalkundig handboek*. Antwerpen: De Sikkel.
- Groot, J. De & S. Heynemann (1973). *Schrijven - en gelezen worden*. Groningen: Tjeenk Willink.
- Haje, C.F. (1932). *Taalschut: schrijf weer Nederlandsch*. Leiden: Leidsche uitgeversmaatschappij.

- Haver, J. Van (1989). *Noorderman en Zuiderman: het taalverdriet van Vlaanderen*. Tielt: Lannoo.
- *Heidbuchel, H. (1967). *Hoe zeg en schrijf ik het?* Hasselt: HeideLand.
- Heidbuchel, H. (1973). *Hoe schrijf ik een keurige brief?* Hasselt: HeideLand-Orbis.
- *Heidbuchel, H. (1975). *Correct Nederlands voor alledag*. Hasselt: HeideLand.
- *Heidbuchel, H. & G.J. Luyten (1985). *Hoe zeg en schrijf ik het?* Antwerpen: HeideLand-Orbis.
- *Hendrickx, R. (2003). *Stijlboek VRT*. Utrecht: Van Dale Lexicografie.
- *Hermkens, H.M. (1966). *Verzorgd Nederlands*. 's Hertogenbosch: Malmberg.
- *Hermkens, H.M. (1972). *Verzorgd Nederlands*. 's Hertogenbosch: Malmberg.
- Hogen, R. Van (1991). *Praktische cursus formuleren*. Groningen: Wolters-Noordhoff.
- Horst, P.J. Van der (1993). *Redactiewijzer: praktische handleiding voor het taalkundig en typografisch verzorgen van teksten*. Den Haag: Sdu.
- *Horst, P.J. Van der (1998). *Redactiewijzer: praktische handleiding voor het taalkundig en typografisch verzorgen van teksten*. Den Haag: Sdu.
- Japin, B. (1969). *Taaltje wèl, taaltje niet: spelen met taal*. Rotterdam: Donker.
- Jonker, DA, P.J. Van den Hoven & H.T.J.F. Van Maarseveen (1983). *Recht spreken, krom schrijven: een bescheiden taalgids voor juristen*. Deventer: Kluwer.
- Kempeneers, P. (1973). *ABN tips*. Antwerpen: De Sikkel.
- *Klein, M & M. Visscher (1992). *Handboek verzorgd Nederlands: spellingregels, stijladviezen*. Groningen: Noordhoff.
- *Klein, M & M. Visscher (1996). *Handboek verzorgd Nederlands: spellingregels, stijladviezen*. Groningen: Noordhoff.
- Lamers, H.A.J.M. (1979). *Hoe schrijf ik een wetenschappelijke tekst? Een handleiding om scripties, verslagen, rapporten en beleidsnota's te schrijven*. Muiderberg: Coutinho.
- Lamers, H.A.J.M. (1991). *Handleiding voor PR- en reclametechnieken*. Muiderberg: Coutinho.
- *Lamers, H.A.J.M. (1997). *Handleiding voor beleidsteksten: een handleiding om beleidsnota's, beleidsbrieven, jaarverslagen en notulen op te stellen*. Bussum: Coutinho.

- Maks, R. & M. De Koning (1987). *Brieven in stijlen en soorten*. Groningen: Wolters-Noordhoff.
- Meert, H. (1897). *Distels: proeve van taalzuivering te gebruiken bij het onderwijs in de Nederlandsche taal*. Brussel: Lebègue.
- Meert, H. & C.H. Peeters (1941). *Onkruid onder de tarwe: proeve van taalzuivering*. Turnhout: Brepols.
- Meijer, A.C. (1904). *Wandelingen op Nederlandsch taalgebied. Bijdragen tot taalzuivering*. Antwerpen: Nederlandsche boekhandel.
- Meijers, J.A. (1952). *Zin en onzin in het hedendaags taalgebruik*. Amsterdam: Van Ditmar.
- Mesters, G.A. (s.d.). *Schrijf/spreek zuiver Nederlands: onderhoudend taalboek voor alle Nederlanders of Hoe men zichzelf een pluim op de hoed steekt door zijn moedertaal te eren*. Utrecht: Spectrum.
- Moerdijk, J.M. & P. Sluimer (1967). *Taalgids voor de ambtenaar*. 's Gravenhage: VUGA.
- Moerdijk, J.M. & P. Sluimer (1976). *Taalgids voor de ambtenaar*. 's Gravenhage: VUGA.
- Moerdijk, J.M. & P. Sluimer (1989). *Taalgids voor de ambtenaar*. Culemborg: Lemma.
- Nederhoed, P. (1983). *Wie wil er even notuleren? Een beknopte handleiding voor het schrijven van vergaderverslagen*. Deventer: Van Loghum Slaterus.
- Nederhoed, P. (1987). *Helder rapporteren: een handleiding voor het schrijven van rapporten, scripties, nota's en artikelen in wetenschap en techniek*. Deventer: Van Loghum Slaterus.
- Nederhoed, P. (1990). *Helder rapporteren: een handleiding voor het schrijven van rapporten, scripties, nota's en artikelen in wetenschap en techniek*. Deventer: Van Loghum Slaterus.
- Nierop, M.J.L. Van (1972). *Honderd taaie belgicismen*. Hasselt: Heideland-Orbis.
- Nierop, M.J.L. Van (1978). *Nog honderd hardnekkige belgicismen: verzameling van taaltuinen en andere bijdragen*. Hasselt: Heideland-Orbis.

- Ommen, H. Van & E. Van Kuppenveld (1986). *Professionele bedrijfscommunicatie: een handboek voor alle tekstschrijvers bij bedrijven en non-profit-organisaties*. Groningen: Wolters-Noordhoff.
- Onrust, M, A. Verhagen & R. Doeve (1993). *Formuleren*. Houten: Bohn Stafleu Van Loghum.
- *Overduin, B. (1986). *Rapporteren. Het schrijven van rapporten, nota's, scripties en artikelen*. Utrecht: Aula.
- *Permentier, L. & L. Van Den Eynden (1998). *Stijlboek*. Groot-Bijgaarden: Scoop.
- *Permentier, L. (2003). *Stijlboek*. Roeselare: Roularta.
- *Pol, J. Van de (1994). *In goed Nederlands: overzichtelijke taalgids met actuele voorbeelden uit de krant; voor op het werk, op school en thuis*. 's Gravenhage: Sdu.
- *Renkema, J. (1979). *Schrijfwijzer: handboek voor duidelijk taalgebruik*. 's Gravenhage: Staatsuitgeverij.
- *Renkema, J. (1995). *Schrijfwijzer*. 's Gravenhage: Sdu.
- Sanders, E. & K. Metselaar (2000). *Stijlboek NRC Handelsblad*. Utrecht: Van Dale Lexicografie.
- *Seghers, A. & A. Vandenheede (1972). *Praktische taalgids: grepen uit de rijkdom van onze taal*. Gent: Norma.
- *Smedts, W. (2000). *Nederlandse taalbeheersing*. Leuven: Acco.
- Smeyers, J., K. Spillebeen & P. Cockx (1996). *Ik schrijf zonder fouten*. Leuven: Davidsfonds / Clauwaert.
- Snoeck Henkemans, A.F. (1990). *Schrijven. Handleiding voor het opstellen van zakelijke teksten*. Leiden: Nijhoff.
- Stassar, P. (1984). *Beheers je taal: taal, schrijven, spreken, solliciteren: een cursus taalbeheersing voor HBO en WO*. Leiden: Nijhoff.
- Sterkenburg, P.G.J. & M.C. Van den Toorn (1997). *Veertig jaar journaal, veertig jaar taal*. Den Haag: Sdu.
- Steen, P. Van (1987). *Rapporteren in organisaties*. Groningen: Wolters-Noordhoff.
- * *Taalbaak* (1994). 61.1.
- **Taalgebruiksadviezen* (1983). Weert: M en P boeken.

- *Talpaert, R., P. Buyse & J.L. Pauwels (1966). *Verzorgde schrijftaal: een praktische handleiding ten gerieve van Nederlands schrijvenden*. Heule: UGA.
- Thomas, J. (1968). *Tekstkritiek en stijlvorming. Behandeling van veel voorkomende stijlfouten*. Gent: Story-Scientia.
- *Tiggeler, E. (2000). *Taalwijzer voor de overheid*. Den Haag: Sdu.
- Tiggeler, E. (2002). *Vraagbaak Nederlands: van spelling tot stijl: snel een helder antwoord op praktijkvragen over taal*. Den Haag: Sdu.
- Timmermans, H. (1948). *Spreken en schrijven wij goed Nederlands?* Antwerpen: Boekuil en Karveel.
- Veering, J. & W. Pée (s.d.). *Spelenderwijs (zuiver) Nederlands: een wegwijzer door het troebele grensland van goed en fout, van wat mag en wat niet mag, van wat allebei mag en wat geen van beide mag*. Amsterdam: Becht.
- Verdenius, A. (1946). *In de Nederlandse taaltuin: wandelingen en waarnemingen*. Amsterdam: Elsevier.
- Vroom, B. (1994). *De tekst getest: handleiding voor het presenteren van schriftelijk materiaal*. Assen: Van Gorcum.
- Wethlij, A. (1985). *Lees- en schrijftechniek*. Antwerpen: Heideland-Orbis.

Appendix B: technische gegevens van de zoekopdrachten

Gegeven de syntactische aard van het studieobject, gegeven de specificiteit van de onderzoekscontext en het ongeannoteerde karakter van het CONDIV-corpus, waaruit een deel van de relevante gegevens gehaald moesten worden, hebben we gebruik gemaakt van de reguliere-expressiesyntaxis (zie bv. Friedl 1998) om de relevante gegevens uit de corpora te halen. In B.1 tot en met B.6 worden de verschillende zoekopdrachten voorgesteld. Dat er meer dan 1 zoekopdracht geformuleerd werd, heeft enerzijds te maken met de specificiteit van de gebruikte corpora (voor het CGN werden aparte reguliere expressies geformuleerd, omdat daar gebruik gemaakt kon worden van de woordsoortannotatie) en anderzijds met de verscheidenheid aan plaatsingsmogelijkheden (enkel de deelwoordspecifieke morfemen *-ge-*, *-be-* en *-ver-* kunnen voorafgegaan worden door een partikel). We bespreken achtereenvolgens de zoekopdrachten voor de consultatie van het CONDIV-corpus en het CGN (R5).

De beide CONDIV-zoekopdrachten (B.1 en B.2) bestaan uit twee grote delen, verbonden door een Boolese OF-knoop: het eerste deel spoort de eindgroepen in de groene volgorde [PART+AUX] op, het tweede deel de eindgroepen in de rode volgorde [AUX+PART]. Zoekopdracht B.1 focust op de gewenste constructie waarvan de deelwoorden de morfemen *-ge-*, *-be-* en *-ver-* bevatten, B.2 focust op de targetconstructie met alle andere deelwoorden (er was een zekere overlap tussen de resultaten van de beide zoekopdrachten; in de zuiveringsoperatie werden de dubbels uit de dataset gehaald). De zoekopdrachten werden uitgevoerd met behulp van *Abundantia Verborum* (Speelman 1997).

De reguliere expressies die gecreëerd werden voor de consultatie van het CGN (R5), B.3 tot en met B.6, en geïntegreerd werden in een Pythonscript, zijn beknopter, enerzijds omdat gebruik gemaakt kon worden van de woordsoortannotatie, maar anderzijds ook omdat om technische redenen minder selectiecriteria formeel gedefinieerd konden worden. De eerste twee CGN-expressies focussen op eindgroepen in de rode [AUX+PART] (B.3) en groene [PART+AUX] (B.4) volgorde met een deelwoord en een hulpwerkwoord, de laatste twee CGN-expressies focussen op eindgroepen in de groene [PART+AUX] (B.5) en rode [AUX+PART] (B.6) volgorde waarvan het deelwoord door de CGN-annotatoren als ‘adjectivisch’ gecategoriseerd

werd. Met de laatste twee zoekopdrachten probeerden we enerzijds de deelwoorden op te zoeken die wegens foute of inconsequente annotatie door de mazen van het net geglipt waren (zie Van Eynde 2004: 26-27 voor de door het CGN gehanteerde criteria) en anderzijds de deelwoorden te traceren die synchroon niet (meer) als deelwoord beschouwd worden, maar wel de formele kenmerken hebben van een deelwoord (daardoor bevatten de CGN-dataset en de CONDIV-dataset vergelijkbaar materiaal: zie paragraaf 2.5 en paragraaf 5.2.4.2). Tot slot moet opgemerkt worden dat het Python-script de lijnen negeert die met <au (met meta-informatie over de spreker) beginnen en niet toelaat dat de targetconstructie doorbroken wordt door een <au-lijn.

B.1 Zoekopdracht 1 in het CONDIV-corpus

```
[a-z]+[^a-z, ?; .: !-]+dat [^a-z, ?; .: !-]+[a-z]+[^, ?; .: !-]*[^a-
z, ?; .: !-]+[a-z]*(ge|be|ver) [a-zëï]+[dtn] [^a-z, ?; .: !-]+, ?; .: !-
]+(ben|bent|is|zijn|was|waren|heb|hebt|heeft|hebben|had|hadden|
word|wordt|worden|werd|werden) [^a-z, ?; .: !'"-]*[^?.:; , !'"-
]*[!..?]
```

OF

```
a-z]+[^a-z, ?; .: !-]+dat [^a-z, ?; .: !-]+[a-z]+[^, ?; .: !-]*[^a-
z, ?; .: !-]+, ?; .: !-
]+(ben|bent|is|zijn|was|waren|heb|hebt|heeft|hebben|had|hadden|
word|wordt|worden|werd|werden) [^a-z, ?; .: !-]+[a-
z]*(ge|be|ver) [a-zëï]+[dtn] [^a-z, ?; .: !'"-]*[^?.:; , !'"-
]*[!..?]
```

B.2 Zoekopdracht 2 in het CONDIV-corpus

```
[a-z]+[^a-z, ?; .: !-]+dat [^a-z, ?; .: !-]+[a-z]+[^, ?; .: !-]*[^a-
z, ?; .: !-
]+(her|ont|aan|achter|om|onder|over|er|mis|vol|voor|door|weer) [
a-zëï]+[dtn] [^a-z, ?; .: !-
]+(ben|bent|is|zijn|was|waren|heb|hebt|heeft|hebben|had|hadden|
word|wordt|worden|werd|werden) [^a-z, ?; .: !'"-]*[^?.:; , !'"-
]*[!..?]
```

OF

```
[a-z]+[^a-z,?;.:!-]+dat[^a-z,?;.:!-]+[a-z]+[^,?;.:!-]*[^a-
z,?;.:!-]+,?;.:!-
]+(ben|bent|is|zijn|was|waren|heb|hebt|heeft|hebben|had|hadden|
word|wordt|worden|werd|werden)[^a-z,?;.:!-
]+(her|ont|aan|achter|om|onder|over|er|mis|vol|voor|door|weer)[
a-zëï]+[dtn][^a-z,?;.:!'"-]*[^?.:;,'!'"-]*[!?.?]
```

B.3 Zoekopdracht 1 in het CGN

```
WW\ (pv\, (tgw|verl)
```

```
WW\ (vd\, vrij\, zonder\)
```

B.4 Zoekopdracht 2 in het CGN

```
WW\ (vd\, vrij\, zonder\)
```

```
WW\ (pv\, (tgw|verl)
```

B.5 Zoekopdracht 3 in het CGN

```
^[a-z]*(ge|be|ver)[a-z]+[dtn]\tADJ\ (vrij\, basis\, zonder\)  

^(ben|bent|is|zijn|zijt|was|waren|heb|hebt|heeft|hebben|had|hadd  

den|word|wordt|worden|werd|werden)\tWW\ (pv\, (tgw|verl)
```

B.6 Zoekopdracht 4 in het CGN

```
^(ben|bent|is|zijn|zijt|was|waren|heb|hebt|heeft|hebben|had|hadd  

den|word|wordt|worden|werd|werden)\tWW\ (pv\, (tgw|verl)  

^[a-z]*(ge|be|ver)[a-z]+[dtn]\tADJ\ (vrij\, basis\, zonder\)
```


Appendix C: criteria ter classificatie van de laatste zinsplaats als complement of adjunct

Appendix C biedt een overzicht van de criteria die in paragraaf 4.4.3.2 gebruikt zijn om een laatste zinsplaats te classificeren als adjunct of als complement. Achtereenvolgens sommen we de criteria op voor een laatste zinsplaats met het hoofdwerkwoord als hoofd (C.1) en voor een laatste zinsplaats met een preverbaal gesitueerd nomen, adjectief of adverbium als hoofd (C.2).

C.1 Het hoofd van de laatste zinsplaats is het hoofdwerkwoord:

- De laatste zinsplaats is het subject van de zin OF
- De laatste zinsplaats is een direct of indirect object van de zin OF
- De laatste zinsplaats is een voorzetselvoorwerp met een vast voorzetsel (detectie van de vaste voorzetsels op basis van ANS 1997 en EGVD 1.4 (2004)) OF
- De laatste zinsplaats is een bepaling van gesteldheid

=> *De laatste zinsplaats is een complement*

- De laatste zinsplaats is een bijwoordelijke bepaling in de zin

=> *De laatste zinsplaats is een adjunct*

C.2 Het hoofd van de laatste zinsplaats is preverbaal gesitueerd:

- De preverbale constituent is een adjectief of bijwoord en de laatste zinsplaats is een comparatieve constituent OF
- De preverbale constituent is een nomen en de laatste zinsplaats is een voorzetselvoorwerp met een vast voorzetsel (detectie van de vaste voorzetsels op basis van ANS 1997 en EGVD 1.4 (2004)) OF
- De preverbale constituent is een nomen en de laatste zinsplaats is een beknopte bijzin met de waarde van een voegwoordelijke bijzin OF

- De preverbale constituent is een nomen en de laatste zinsplaats is een bijzin ingeleid door het grammaticaal verbindend voegwoord *dat* of *of* of een vraagwoord.

=> *De laatste zinsplaats is een complement*

- De preverbale constituent is een adjectief en de laatste zinsplaats verwijst naar de oorzaak of de oorsprong van de preverbale constituent OF
- De preverbale constituent is een adjectief en verwijst zowel naar de laatste zinsplaats als naar het subject of een object van de zin OF
- De preverbale constituent is een bijwoord en de laatste zinsplaats drukt het gevolg uit OF
- De preverbale constituent is een nomen en de laatste zinsplaats is een beknopte bijzin die gebruikt wordt als relatiefzin.
- De preverbale constituent is een nomen en de laatste zinsplaats is een relatiefzin.
- De preverbale constituent is een nomen en de laatste zinsplaats is een comparatieve constituent.
- De preverbale constituent is een nomen en de laatste zinsplaats is een voorzetselvoorwerp zonder vast voorzetsel.

=> *De laatste zinsplaats is een adjunct*

Appendix D: overzicht van de items in de disambigueringsenquête

In appendix D geven we een overzicht van de testitems die gebruikt zijn om uit te maken welke contextuele criteria de categoriale status van het deelwoord disambigueren (voor een beschrijving van de enquête, zie paragraaf 5.2.4.2: module 3, submodule A).

Deel 1 (rood:groen = 18:24):

1. Beschuldigingen dat hij in 1963 betrokken was bij een aanval op een synagoge in Buenos Aires droegen tot zijn vertrek bij.
2. bleek dat de gesmokkelde diamanten gestolen waren uit de diamantmijnen in Kimberley en aan de Westkust.
3. Gezien dat grote stukken landbouwgrond en infrastructuurwerken zijn verwoest
4. jaar dat de laatste gestraften dood en begraven zijn?
5. uit dat ze al gebruikt zijn.
6. zeggen dat als een gevolg daarvan het doorbuigen niet alleen is verminderd.
7. Beweert dat de wapens bij hem gedeponeerd zijn door een stadsgenoot.
8. Is dat de gemeente verplicht is jaarlijks de eventuele tekorten van de kerkfabrieken te dekken (art. 131,9° van de gemeentewet).
9. maakt dat de campus nu vooral bezet is door lieden die in de hoofdstad op congres zijn.
10. Houden dat zijn bezittingen in de VS door Washington bevroren waren.
11. ore dat een telefoontoestel huren bij Belgacom nu helemaal niet meer verplicht is.
12. toejuichen dat er zo'n beslissing is genomen.
13. zegt dat de rellen van 27 juli in Jakarta waren georganiseerd door linkse Indonesiërs die het bewind van Suharto omver willen werpen.
14. aan dat de Tijgers verre van verslagen zijn.
15. vrezen dat de toeschouwers de indruk krijgen dat het stuk op waargebeurde feiten gebaseerd is.
16. beweren dat onze huidige regeerders niet verwittigd waren.

17. dan dat het vertrouwen van veel klanten in de Luxemburgse banksector eventjes geschokt is en dat er rond de Kredietbank Luxembourg en - onvermijdelijk - haar Belgische zusterbank een verdacht geurtje blijft hangen.
18. op dat het nationale karakter van de Duitsers na de oorlog is veranderd.
19. meent dat Duitsland na 1945 opnieuw is gevormd naar westerse waarden.
20. rapport dat vele Turkse artsen ervan overtuigd zijn dat iedereen die in Turkije wordt gearresteerd ook wordt gefolterd.
21. doorschemeren dat de spullen gestolen zijn en daardoor goedkoop van de hand aan.
22. vraaggesprek dat er gepoogd was zijn reputatie te breken.
23. vaststellen dat het eigendomsrecht over hun gronden totaal uitgehold is door de erfdiensbaarheden van openbaar nut die de Vlaamse overheid kwistig oplegt.
24. regelmatig dat de verkiezingsoverwinning verre van verworven is.
25. om dat de weg was afgesloten wegens het noodweer.
26. echter dat de abortus op een van de twee foetussen al een maand geleden was uitgevoerd.
27. erop dat die kloof sedertdien nog vergroot is.
28. zo dat de situatie in Sarajevo fundamenteel veranderd is.
29. geweten dat de verkiezingscampagne was gefinancierd met drugsgelden van het cocaïnekartel van Cali.
30. best dat het regime in Zaïre niet fundamenteel is veranderd.
31. zeggen dat 90 procent van alle bergcampings in Spanje aan een rivier zijn gebouwd.
32. beseffen dat onze liefde op een vooroordeel is gestoeld.
33. daaruit dat de alpenweide in Romeinse tijd intensief is gebruikt.
34. meent dat de rechtbank verplicht is die informatie te verstrekken.
35. op dat de Vlaamse arbeidersbeweging veel minder goed was georganiseerd dan de Waalse.
36. dat Angelo Movio niet uitgenodigd was op de herdenkingsceremonie.
37. me dat uw verslaggever de lezers deed inzien dat Angelo Movio onrecht is aangedaan.
38. het dat nog niet alle geldtransporten zijn uitgerust met een systeem dat de lading waardeloos maakt of vernietigt bij een overval.

39. vermoeden dat de elektriciteit was afgesloten in het pand dat officieel niet bewoond was.
40. blijkt dat de meningsverschillen nog niet helemaal zijn bijgelegd.
41. toegegeven dat deze illegalen een kalmeermiddel was gegeven om incidenten te vermijden tijdens hun gedwongen terugkeer.
42. dat alle leden van de groep die onder de Franse wetgeving vallen (onder meer de ouders van in Frankrijk geboren en dus Franse kinderen) op 26 juni gelegaliseerd zijn.

Deel 2 (rood:groen = 18:23):

1. is dat een demo van ons op Studio Brussel is beland en daar aansloeg.
2. Zei dat door haar hoofd was geflitst dat haar lichaam na een eventueel ongeluk geïdentificeerd moest kunnen worden en dat zij uit voorzorg haar rijbewijs in haar beha had gestopt.
3. toe dat recent ook in het KIB de invloed van de PKK is toegenomen.
4. uitwijzen dat België hiermee in een doodlopend straatje beland is.
5. ongetwijfeld dat er een lichting pril talent opgestaan is.
6. toegeven dat Dudok daar op bijzondere wijze in geslaagd is.
7. bedrijfsfilosofie dat we tot afvalverbranding zijn gekomen.
8. zei dat hij niet op de bank was geraakt en later zou betalen.
9. vergeten" dat eerder deze eeuw de rivier verscheidene keren buiten zijn oevers was getreden.
10. blijkt dat de verhuur van appartementen en villa's aan de kust in de eerste helft van het jaar is teruggelopen met 10 procent.
11. verstaan dat de strijdmacht eerder al tot de conclusie was gekomen dat er geen fouten zijn begaan in Ruanda.
12. zondag dat Bossi een avontuurlijke weg is ingeslagen en dat hij moet oppassen dat hij de justitie niet over de vloer krijgt.
13. niet dat de volksdevotie in Vlaanderen is verdwenen.
14. ontkende dat hij overgelopen was.
15. voorstellen dat er ooit een kunstenaar met een verfkwast tegenaan was gegaan.
16. is dat er inderdaad een einde aan de beproevingen en de tirannie is gekomen.

17. in 1986 dat vanuit Vlaanderen nà 1945 een half miljoen migranten naar Wallonië is getrokken.
18. beide dat de beschermwaardigheid van het ongeboren leven sterk is afgenomen.
19. aan dat hij omgekomen is.
20. zijn dat we na WO II in een overgangsfase naar een nieuw tijdperk zijn aanbeland.
21. mee dat de IJzerbedevaart meer en meer in het vaarwater van politieke partijen is terechtgekomen?
22. op dat de reis wel degelijk is afgelopen.
23. bestaat dat de meisjes in een prostitutienetwerk zijn beland.

Appendix E: overzicht van de deelwoorden in de drie directe-verbaliteitslijsten

In appendix E geven we een overzicht van de deelwoordtypes in onze dataset waarvan op basis van de directe-verbaliteitsmeting vastgesteld kon worden dat ze als sterk werkwoordelijk opgevat moeten worden. In de lijst DVL 1 worden de deelwoorden opgesomd die in CGN (R5) en Celex dubbel zo vaak als werkwoord geannoteerd werden dan als adjectief. DVL 2 biedt een overzicht van de deelwoorden die drie keer zo vaak als werkwoord geannoteerd werden dan als adjectief. DVL 3 geeft een overzicht van de deelwoorden die enkel als werkwoord geannoteerd werden en nooit als adjectief (zie paragraaf 5.2.4.2: module 3, submodule B voor meer uitleg).

DVL 1 \geq 0.66 (n = 194)

aangehouden, aangemoedigd, aangetast, afgehaald, afgeleverd, afgenomen, afgepakt, afgestemd, afgeweken, bedoeld, bedreigd, begeleid, begonnen, begraven, beland, belegd, benoemd, bereikt, beschouwd, beschuldigd, besteed, bestemd, bestraft, bestreden, betekend, bevallen, bevestigd, bewezen, bijgekomen, binnengebracht, doodgeschoten, doordrongen, geaccepteerd, gebaseerd, gebeurd, gebracht, gebruikt, gedaald, gedoemd, gedood, gefaald, gefilmd, gefinancierd, gefloten, gegaan, gegroeid, gehaald, gehoord, gekeerd, gekend, geknoeid, gekomen, gekozen, gekregen, gekropen, gelegd, gelopen, genegeerd, genoodzaakt, gepland, gepoogd, geproduceerd, gereden, geschonken, geschorst, geschrokken, gesneden, gesneuveld, gespaard, gesproken, gestapt, gestart, gestegen, gesteund, gestolen, gestoten, gestreden, gestuurd, getrapt, getreden, gevochten, gevolgd, gevonden, gewaarschuwd, gewerkt, goedgekeurd, hersteld, ingebouwd, ingediend, ingegaan, ingehuurd, ingelicht, ingeslagen, meegegaan, meegenomen, misbruikt, misgelopen, mislukt, omgekomen, omschreven, ondervraagd, onderwezen, onderzocht, ontdekt, ontgaan, ontnomen, ontploft, ontslagen, ontstaan, ontvangen, ontvoerd, opgedoken, opgegaan, opgegraven, opgegroeid, opgelopen, opgelost, opgenomen, opgepakt, opgeschreven, opgeslagen, opgesloten, opgestaan, opgesteld, opgetreden, overeengekomen, overgegaan, overkomen, overschreden, stilgevallen, terechtgekomen, teruggekeerd, teruggelopen, teruggevonden, thuisgekomen, toegelaten, toegenomen, toegestaan, toevertrouwd, uitbetaald, uitgegroeid, uitgehold, uitgelegd, uitgelopen, uitgenodigd, uitgeoefend, uitgeroeid, uitgeschakeld, uitgeschreven, uitgespeeld, uitgevoerd, veranderd, verbeterd, verbonden, verdeeld, verdwenen, vereist, vergroot, verkocht, verkozen, verkracht, verleend, vermeld, verminderd, vermoord, vernomen, veroordeeld, veroorzaakt, verschenen, versterkt, vertaald, vertegenwoordigd, vertrokken, vervuld, verwekt, verwittigd, verworven, verzoend, verzonnen, voorbereid, voorbijgestreefd, vrijgelaten, weggelegd, weggelopen, gelaten, ervaren, gevallen, geweest, geworden, gebleven- overgestoken, geboekt, gestorven, gearresteerd

DVL 2 \geq 0.75 (n = 191)

aangehouden, aangemoedigd, aangetast, afgehaald, afgeleverd, afgenomen, afgepakt, afgestemd, afgeweken, bedoeld, bedreigd, begeleid, begonnen, begraven, beland, belegd, benoemd, bereikt, beschouwd, beschuldigd, besteed, bestemd, bestraft, bestreden, betekend, bevallen, bevestigd, bewezen, bijgekomen, binnengebracht, doodgeschoten, doordrongen, geaccepteerd, gebaseerd, gebeurd, gebracht, gebruikt, gedaald, gedoemd, gedood, gefaald, gefilmd, gefinancierd, gefloten, gegaan, gegroeid, gehaald, gehoord, gekeerd, gekend, geknoeid, gekomen, gekozen, gekregen, gekropen, gelegd, gelopen, genegeerd, genoodzaakt, gepland, gepoogd, geproduceerd, gereden, geschonken, geschorst, geschrokken, gesneden, gesneuveld, gespaard, gesproken, gestapt, gestart, gestegen, gesteund, gestolen, gestoten, gestreden, gestuurd, getrapt, getreden, gevochten, gevolgd, gevonden, gewaarschuwd, gewerkt, goedgekeurd, hersteld, ingebouwd, ingediend, ingegaan, ingehuurd, ingelicht, ingeslagen, meegegaan, meegenomen, misbruikt, misgelopen, mislukt, omgekomen, omschreven, ondervraagd, onderwezen, onderzocht, ontdekt, ontgaan, ontnomen, ontploft, ontslagen, ontstaan, ontvangen, ontvoerd, opgedoken, opgegaan, opgegraven, opgegroeid, opgelopen, opgelost, opgenomen, opgepakt, opgeschreven, opgeslagen, opgesloten, opgestaan, opgesteld, opgetreden, overeengekomen, overgegaan, overkomen, overschreden, stilgevallen, terechtgekomen, teruggekeerd, teruggelopen, teruggevonden, thuisgekomen, toegelaten, toegenomen, toegestaan, toevertrouwd, uitbetaald, uitgegroeid, uitgehold, uitgelegd, uitgelopen, uitgenodigd, uitgeoefend, uitgeroeid, uitgeschakeld, uitgeschreven, uitgespeeld, uitgevoerd, veranderd, verbeterd, verbonden, verdeeld, verdwenen, vereist, vergroot, verkocht, verkozen, verkracht, verleend, vermeld, verminderd, vermoord, vernomen, veroordeeld, veroorzaakt, verschenen, versterkt, vertaald, vertegenwoordigd, vertrokken, vervuld, verwekt, verwittigd, verworven, verzoend, verzonnen, voorbereid, voorbijgestreefd, vrijgelaten, weggelegd, weggelopen, geweest, geworden, gebleven, overgestoken, geboekt, gestorven, gearresteerd

DVL 3 = 1 (n = 190)

aangehouden, aangemoedigd, aangetast, afgehaald, afgeleverd, afgenomen, afgepakt, afgestemd, afgeweken, bedoeld, bedreigd, begeleid, begonnen, begraven, beland, belegd, benoemd, bereikt, beschouwd, beschuldigd, besteed, bestemd, bestraft, bestreden, betekend, bevallen, bevestigd, bewezen, bijgekomen, binnengebracht, doodgeschoten, doordrongen, geaccepteerd, gebaseerd, gebeurd, gebracht, gebruikt, gedaald, gedoemd, gedood, gefaald, gefilmd, gefinancierd, gefloten, gegaan, gegroeid, gehaald, gehoord, gekeerd, gekend, geknoeid, gekomen, gekozen, gekregen, gekropen, gelegd, gelopen, genegeerd, genoodzaakt, gepland, gepoogd, geproduceerd, gereden, geschonken, geschorst, geschrokken, gesneden, gesneuveld, gespaard, gesproken, gestapt, gestart, gestegen, gestolen, gestoten, gestreden, gestuurd, getrapt, getreden, gevochten, gevolgd, gevonden, gewaarschuwd, gewerkt, goedgekeurd, hersteld, ingebouwd, ingediend, ingegaan, ingehuurd, ingelicht, ingeslagen, meegegaan, meegenomen, misbruikt, misgelopen, mislukt, omgekomen, omschreven, ondervraagd, onderwezen, onderzocht, ontdekt, ontgaan, ontnomen, ontploft, ontslagen, ontstaan, ontvangen, ontvoerd, opgedoken, opgegaan, opgegraven, opgegroeid, opgelopen, opgelost, opgenomen, opgepakt, opgeschreven, opgeslagen, opgesloten, opgestaan, opgesteld, opgetreden, overeengekomen, overgegaan, overkomen, overschreden, stilgevallen, terechtgekomen, teruggekeerd, teruggelopen, teruggevonden, thuisgekomen, toegelaten, toegenomen, toegestaan, toevertrouwd, uitbetaald, uitgegroeid, uitgehold, uitgelegd, uitgelopen, uitgenodigd, uitgeoefend, uitgeroeid, uitgeschakeld, uitgeschreven, uitgespeeld, uitgevoerd, veranderd, verbeterd, verbonden, verdeeld, verdwenen, vereist, vergroot, verkocht, verkozen, verkracht, verleend, vermeld, verminderd, vermoord, vernomen, veroordeeld, veroorzaakt, verschenen, versterkt, vertaald, vertegenwoordigd, vertrokken, vervuld, verwekt, verwittigd, verworven, verzoend, verzonnen, voorbereid, voorbijgestreefd, vrijgelaten, weggelegd, weggelopen, geweest, geworden, gebleven, overgestoken, geboekt, gestorven, gearresteerd

Appendix F: lijst van de deelwoorden die adjectieftolerant zijn

Appendix F geeft een overzicht van de deelwoordtypes in onze dataset die als adjectieftolerant of adjectiefcompatibel beschouwd kunnen worden (zie paragraaf 5.2.4.2: module 3, submodule C). Achtereenvolgens komt de lijst met deelwoordtypes die in CGN (R5) minstens één keer:

- in preminale positie voorkomen;
- met een comparatief- of superlatiefsuffix voorkomen;
- in een omschreven comparatief of superlatief gebruikt worden;
- gegradeerd worden;
- met *on-* geprefigeerd worden;
- gesubstantiveerd worden.

De deelwoordtypes in vetjes komen minstens 20 keer voor in de genoemde morfosyntactische context (bv. *afgelopen* komt minstens 20 keer voor in preminale positie).

Deelwoorden in preminale positie; n =210

aangebracht, aangekomen, aangepast, aangesloten, aangetast, aangewezen, **afgelopen**, afgenomen, afgerond, afgesloten, afgesproken, afgestemd, afgeweken, afgewerkt, bedoeld, bedreigd, bedwongen, begeleid, begraven, **bekend**, belegd, **bepaald**, **beperkt**, bereikt, beschadigd, beslist, besloten, **besmet**, **betaald**, **betrokken**, bevroren, bewapend, bewezen, bezet, bezorgd, binnengebracht, ervaren, geaccepteerd, geassimileerd, gebaseerd, gebeurd, gebleken, geblust, gebonden, gebouwd, gebracht, **gebroken**, gebruikt, geconcentreerd, gecremeerd, gedaan, gedekt, gedeprimeerd, gedoemd, gedreven, gedrukt, gefinancierd, gefolterde, **gegeven**, gegroeid, gegrond, geheeld, gehoord, gehuwd, gekend, gekomen, gekopieerd, gekozen, gelaten, geleerd, gelegaliseerd, gelegd, gelokaliseerd, gemaakt, genegeerd, genomen, geopend, georganiseerd, georkestreerd, geplaatst, gepland, geplunderd, geproduceerd, geraakt, gerechtvaardigd, gereden, geregeld, **gericht**, gescheiden, **geschikt**, geschonden, geschorst, geslaagd, geslagen, **gesloten**, gesneden, gespaard, gesproken, gestart, gestegen, gesteld, gestolen, gestuurd, getint, getroffen, getrokken, getrouwd, gevallen, gevestigd, gevoerd, gevormd, gewezen, gewijzigd, gewonnen, gezegend, gezet, gezien, gezocht, gezonken, gezwollen, goedgekeurd, hersteld, ingediend, ingelicht, ingenomen, ingeschreven, ingeslagen, ingezet, misbruikt, **mislukt**, omgekomen, ondervraagd, ontdekt, ontgoocheld, ontkoppeld, ontsnapt, ontstaan, ontvangen, opgeblazen, opgelost, opgenomen, opgepakt, opgeschreven, opgeslagen, opgesteld, overdreven, overeengekomen, overleden, overtuigd, samengesteld, stilgevallen, teruggevonden, toegenomen, toegepast, **toegevoegd**, toegewijd, toevertrouwd, uitgelegd, uitgelopen, uitgeput, uitgerust, uitgeschreven, uitgespeeld, uitgesproken, uitgeteld, uitgevoerd, veranderd, **verantwoord**, verbeterd, verboden, verbonden, verdeeld, verdraaid, verdwenen, vereist, vergeten, vergroot, **verkeerd**, verkocht, verkozen, verlopen, **verloren**, vermeld, verminderd, vermoeid, vermoord, veroordeeld, **verplicht**, verschenen, verslagen, versterkt, vertekend, vertrokken, **vertrouwd**, vervalst, vervlogen, verward, verwoest, verworven, verzadigd, verzekerd, verzonnen, voorbereid, vooringenomen, **voortgezet**, voorzien, weggelopen, geweest, geworden, gestorven

Deelwoorden met comparatief/superlatiefsuffix; n = 1

bekend

Deelwoorden met omschreven comparatief/superlatief; n = 84

aangepast, aangewezen, afgenomen, afgesloten, afgesproken, bedoeld, begeleid, begonnen, bekend, bepaald, bereikt, betaald, bevroren, bewezen, bezorgd, ervaren, gebeurd, gebonden, gebouwd, gebroken, gebruikt, gedaan, gedrukt, gefinancierd, gegaan, gegeven, gehaald, gehoord, gekend, gekomen, gekozen, gekregen, geleerd, gemaakt, genegen, geneigd, genomen, geopend, gepland, geregeld, gericht, geschikt, geslaagd, geslagen, gesproken, getrokken, gevochten, gevoerd, gevonden, gewerkt, gewijzigd, gewonnen, gezegd, gezien, hersteld, onderzocht, ontslagen, opgegroeid, opgesteld, opgewassen, overtuigd, teruggehaald, teruggekeerd, teruggevonden, toegepast, toegevoegd, uitgenodigd, uitgerust, uitgesloten, uitgevoerd, veranderd, verbonden, vergeten, verkeerd, verkocht, verplicht, vertrouwd, verzekerd, voldaan, voorzien -geweest, geworden, geconcentreerd, gesteund

Gegradeerde deelwoorden; n = 35

beperkt, uitgesproken, vertrouwd, bezorgd, gelaten, **bekend**, ontgoocheld, gekend, gekende, veranderd, aangedaan, ervaren, aangepast, beperkte, gericht, geschikt, geslaagde, overdreven, verkeerd, gesloten, gegroeid, opgeblazen, ingenomen, toegewijd, onderworpen, gebaseerd, geregeld, gerechtvaardigd, bekende, uitgelegd, geneigd, verbonden, verantwoord, geconcentreerd, gelokaliseerd

Deelwoorden met on-prefix; n = 43

aangedaan, aangepast, bedoeld, begonnen, **bekend**, bepaalde, beperkt, beschadigd, besliste, bestemd, betaald, bezorgd, ervaren, gebonden, geborgen, gebruikt, gedaan, gedekt, gegrond, gehoord, gehuwd, gekend, geopend, geschikt, geschonden, gesteld, getrouwd, gewijzigd, gezien, opgelost, opgemaakt, uitgenodigd, uitgesproken, veranderd, verantwoord, vermeld, verminderd, vervalste, verzekerd, voldaan, voorbereid, voorzien

Gesubstantiveerde deelwoorden; n = 29

bekend, bepaald, beschuldigd, **betrokken**, geborgen, **gegeven**, gekozen, geleerd, geregeld, geschikt, geslaagd, gevallen, gevestigd, gezegd, gezien, ontglipt, overleden, toegepast, toegevoegd, verboden, vereist, verkeerd, verkozen, vermeld, **vermoeden**, verslaafd, verslagen, vertrouwd, verwoest, geoorloofd

Appendix G: puzzelexperiment

Appendix G bevat het testformulier van het puzzelexperiment dat in paragraaf 6.2.3.4 besproken wordt. De stimulusvariabelen zijn:

- Afhankelijke variabele: woordvolgorde in tweeledige werkwoordelijke eindgroepen met *hebben* als hulpwerkwoord en een verleden deelwoord in de vorm van *ge-X-d/t*.
- Onafhankelijke variabelen: lengte van de laatste zinsplaats (0 woorden vs. 3 woorden vs. 6 woorden), structurele status van de laatste zinsplaats (complement van het hoofdwerkwoord vs. adjunct van het hoofdwerkwoord, allebei in de vorm van een voorzetselconstituent).

Het experiment werd in twee versies aangeboden. De versies verschilden enkel van elkaar in de ordening van de experimentele items en niet in de ordening van de zinsstukken: het eerste experimentele item van de ene versie is het laatste in de andere versie, het tweede item in de ene versie is het voorlaatste in de andere, en zo verder. Hieronder is slechts een van de versies opgenomen. De items met een A-code zijn de relevante testitems, die met een B-code zijn fillers. Alvorens het experiment van start ging, werd uitgebreid stilgestaan bij de eerste pagina van het formulier en werd gezegd dat de testzinnen uit de wetenschapsbijlage van een kwaliteitskrant gehaald zijn. Door die context op te geven, probeerden we mogelijke registervariatie in te dijken.

KATHOLIEKE UNIVERSITEIT LEUVEN
 DEPARTEMENT LINGUIÏSTIEK
 BLIJDE-INKOMSTSTRAAT 21
 POSTBUS 33
 B-3000 LEUVEN

Bedankt voor uw deelname aan dit experiment. Hieronder vindt u een aantal zinnen uit de wetenschappsbijlage van een kwaliteitskrant. We hebben die zinnen in verschillende stukken geknipt en vervolgens door elkaar gehaspeld. Uw opdracht bestaat erin om die door elkaar gehaspelde stukken (woorden en woordgroepen) in een volgorde te zetten die u spontaan zou kiezen. U krijgt aan het begin van elke testzin de inleidende basiszin *Het is een feit dat*. U maakt vervolgens de zin af met de stukken die u ter beschikking staan. Bij elk stuk staat ook een nummer. Schrijf bij voorkeur eerst de testzin uit zonder rekening te houden met dat nummer. Zoek daarna het nummer van elk stuk op en vul het nummer in op de daartoe voorziene plaats. Plaats de nummers zo dat het duidelijk is welke volgorde u aan de verschillende stukken gegeven hebt. Een voorbeeldje ter verduidelijking:

	1	2	3
Opgave:	Het is een feit dat er / een bom / gisteravond / ontplofte /		
	Het is een feit dat er.....		
Code			

Stap 1: Zet de zinnen in de door u spontaan gekozen volgorde:

Bijvoorbeeld: *Het is een feit dat er / gisteravond / een bom / ontplofte /*

Stap 2: Zoek de nummers van de constituenten op:

gisteravond komt overeen met nummer 2
een bom komt overeen met nummer 1
ontplofte komt overeen met nummer 3

Stap 3: Zet de nummers in het codevakje:

Code	2	1	3
------	---	---	---

Tijdens het uitvoeren van de taak, moet u de volgende zaken goed in het oog houden:

Voer de test met de nodige concentratie uit, maar blijf niet te lang stilstaan bij de testzinnen: we zijn nl. geïnteresseerd in uw intuïtieve reactie.

Voer de opdracht zin voor zin uit en verander achteraf niets aan voorgaande zinnen.

Voer de opdrachten bladzijde voor bladzijde uit: blader dus niet voor- of achteruit.

Schrijf duidelijk leesbaar.

Alvast bedankt voor uw medewerking.

1 2 3 4
 Het is een feit dat er \ een opvallende overgang \ in de soundtrack \ verleden jaar \ zat \
 Het is een feit dat er

.....
 B15 - Code

--	--	--	--

1 2 3 4 5
 Het is een feit dat hij \ heeft \ na de melding \ geopend \ van het voorval \ de noodlijn \
 Het is een feit dat hij

.....
 A20 - Code

--	--	--	--	--

1 2 3 4
 Het is een feit dat hij \ gezocht \ het verband \ na de botsing \ heeft \
 Het is een feit dat hij

.....
 A12 - Code

--	--	--	--

1 2 3 4
 Het is een feit dat hij \ het onrecht \ geregeld \ met de bezetter \ heeft \
 Het is een feit dat hij

.....
 A2 - Code

--	--	--	--

1 2 3 4
 Het is een feit dat hij \ heeft \ gevraagd \ aan de tandarts \ het potlood \
 Het is een feit dat
 hij.....

.....
 A4 - Code

--	--	--	--

1 2 3 4
 Het is een feit dat hij \ heeft \ de boomhut \ gedeeld \ met het tweetal \
 Het is een feit dat hij

.....
 A1 - Code

--	--	--	--

1 2 3 4
 Het is een feit dat er \ eergisteren \ in de verpakking \ zat \ een scherpe glassplinter \
 Het is een feit dat er

.....
 B2 - Code

--	--	--	--

1 2 3 4
 Het is een feit dat er \ zat \ in het toneelstuk \ een akelig skelet \ bij de laatste vertoning \
 Het is een feit dat er

.....
 B12- Code

--	--	--	--

1 2 3 4
 Het is een feit dat \ stond \ nog een naakte dame \ gisteren \ op de webpagina \
 Het is een feit dat

.....
 B20 – Code

--	--	--	--

1 2 3 4
 Het is een feit dat er \ in de tuin \ stond \ afgelopen week \ een reusachtige distel \
 Het is een feit dat er

.....
 B4 – Code

--	--	--	--

1 2 3
 Het is een feit dat hij \ getekend \ de grondwet \ heeft \
 Het is een feit dat hij

.....
 A22 – Code

--	--	--

1 2 3 4
 Het is een feit dat hij \ de erfenis \ gedeeld \ met de werkloze \ heeft \
 Het is een feit dat hij

.....
 A5 – Code

--	--	--	--

1 2 3 4
 Het is een feit dat er \ zat \ vorige keer \ in het draaiboek \ een dampende revolver \
 Het is een feit dat
 er.....

B11 – Code

--	--	--	--

1 2 3 4 5
 Het is een feit dat hij \ na de aanvang \ het woordje \ heeft \ van het concert \ gezegd \
 Het is een feit dat hij

.....
 A19 – Code

--	--	--	--	--

1 2 3 4
 Het is een feit dat hij \ heeft \ getekend \ na de aankoop \ de grondwet \
 Het is een feit dat hij

.....
 A11 – Code

--	--	--	--

1 2 3 4
 Het is een feit dat hij \ gevraagd \ de subsidie \ heeft \ aan de brouwer \
 Het is een feit dat hij

.....
 A3 – Code

--	--	--	--

1 2 3 4 5
 Het is een feit dat hij \ aan de raadsman \ van de koningin \ heeft \ gevraagd \ de benaming \
 Het is een feit dat hij

.....
 A8 – Code

--	--	--	--	--

1 2 3 4
 Het is een feit dat \ zat \ vorige zondag \ een rode lippenstift \ in de beautycase \
 Het is een feit
 dat.....

B9 – Code

--	--	--	--

1 2 3 4
 Het is een feit dat \ stond \ vorige maand \ in het overzicht \ een ontzettende misser \
 Het is een feit dat

.....
 B16 – Code

--	--	--	--

1 2 3 4
 Het is een feit dat er \ in het water \ zat \ vorige zomer \ een hoog stikstofgehalte \
 Het is een feit dat er

.....
 B3 – Code

--	--	--	--

1 2 3 4
 Het is een feit dat \ lag \ een grijze vissersboot \ vorige week \ in het droogdok \
 Het is een feit dat

.....
 B6 – Code

--	--	--	--

1 2 3 4
 Het is een feit dat hij \ de overgave \ na het voorval \ heeft \ getekend \
 Het is een feit dat hij

.....
 A15 – Code

--	--	--	--

1 2 3 4
 Het is een feit dat \ stond \ vanochtend \ een spijtige vergissing \ in de mededeling \
 Het is een feit dat

.....
 B18 – Code

--	--	--	--

1 2 3 4 5
 Het is een feit dat hij \ gestoord \ de koningin \ van het verraad \ heeft \ na de melding \
 Het is een feit dat hij

.....
 A16 – Code

--	--	--	--	--

1 2 3 4 5
 Het is een feit dat hij \ na de toename \ de inspraak \ heeft \ gestopt \ van het onrecht \
 Het is een feit dat hij

.....
 A18 – Code

--	--	--	--	--

1 2 3 4
 Het is een feit dat \ zat \ een vinnige gitaarsolo \ in het bisnummer \ gisterenavond \
 Het is een feit dat

.....
 B17 – Code

--	--	--	--

1 2 3 4 5
 Het is een feit dat hij \ het vereiste \ aan de adviseur \ gevraagd \ van de zakenman \ heeft \
 Het is een feit dat hij

.....
 A9 – Code

--	--	--	--	--

1 2 3
 Het is een feit dat hij \ het verhoor \ gestoord \ heeft \
 Het is een feit dat hij

.....
 A21 – Code

--	--	--

1 2 3 4 5
 Het is een feit dat hij \ van de fractie \ het vereiste \ geregeld \ heeft \ met de adviseur \
 Het is een feit dat hij

.....
 A7 – Code

--	--	--	--	--

1 2 3 4 5
 Het is een feit dat hij \ de eigendom \ gedeeld \ met het vrouwtje \ van de werkloze \ heeft \
 Het is een feit dat hij

.....
 A6 – Code

--	--	--	--	--

1 2 3 4
 Het is een feit dat er \ vorige zomer \ in de moderubriek \ een reclamefolder \ zat \
 Het is een feit dat er

.....
 B13 – Code

--	--	--	--

1 2 3 4 5
 Het is een feit dat hij \ heeft \ na de stijging \ de subsidie \ van het pensioen \ getekend \
 Het is een feit dat hij

.....
 A17 – Code

--	--	--	--	--

1 2 3 4
 Het is een feit dat er \ in de horrorfilm \ een luguber doodshoofd \ zat \ eergisteren \
 Het is een feit dat er

.....
 B14 – Code

--	--	--	--

1 2 3 4
 Het is een feit dat er \ in de tunnel \ een ernstig ongeval \ gisterenavond \ plaatsvond \
 Het is een feit dat er

.....
 B1 – Code

--	--	--	--

1 2 3 4
 Het is een feit dat hij \ na de sanctie \ de advocaat \ gevraagd \ heeft \
 Het is een feit dat hij

.....
 A14 – Code

--	--	--	--

1 2 3
 Het is een feit dat hij \ heeft \ het fragment \ gezocht \
 Het is een feit dat hij

.....
 A23 – Code

--	--	--

1 2 3
 Het is een feit dat hij \ heeft \ het bouwwerk \ geopend \
 Het is een feit dat hij

.....
 A24 – Code

--	--	--

1 2 3
 Het is een feit dat hij \ de subsidie \ gestopt \ heeft \
 Het is een feit dat hij

.....
 A25 – Code

--	--	--

1 2 3 4 5
 Het is een feit dat hij \ van de notaris \ de schotel \ met de cliënte \ heeft \ gedeeld \
 Het is een feit dat
 hij.....

.....
 A10 – Code

--	--	--	--	--

1 2 3 4
 Het is een feit dat \ een kleurrijke verjaardagskaart \ vorige maand \ zat \ in de brievenbus \
 Het is een feit dat

.....
 B10 – Code

--	--	--	--

1 2 3 4

Het is een feit dat \ in het uitstalraam \ onlangs \ lag \ een dure armband \

Het is een feit dat

.....

B8 – Code

--	--	--	--

1 2 3 4

Het is een feit dat er \ stond \ op dat ogenblik \ in de gang \ een bejaarde schoonmaker \

Het is een feit dat er

.....

B5 – Code

--	--	--	--

1 2 3 4

Het is een feit dat \ zat \ een duidelijke hint \ vorige keer \ in het kortverhaal \

Het is een feit dat

.....

B19 – Code

--	--	--	--

1 2 3 4

Het is een feit dat \ op dat ogenblik \ lag \ in het tabernakel \ een zilveren kruisbeeld \

Het is een feit dat

.....

B7 - Code

--	--	--	--

1 2 3 4

Het is een feit dat hij \ de subsidie \ gedeeld \ na de ingreep \ heeft \

Het is een feit dat hij

.....

A13 - Code

--	--	--	--

Literatuur

- Abney, S. (1996). "Statistical methods and linguistics". In: J. Klavans & P. Resnik (red.), *The balancing act: combining symbolic and statistical approaches to language*. Cambridge: MIT, 1-26.
- Ackerman, F. & A. Goldberg (1996). "Constraints on Adjectival Past Participles". In A.E. Goldberg (red.), *Conceptual structure, discourse and language*. Stanford: CSLI.
- Agresti, A. (1996). *An introduction to categorical data analysis*. New York / Chichester / Brisbane / Toronto / Singapore: Wiley.
- ANS – *Algemene Nederlandse Spraakkunst* (1984). G. Geerts, W. Haeseryn, J.de Rooij, M.C. Van den Toorn (red.). Groningen / Leuven: Wolters-Noordhoff / Wolters.
- ANS – *Algemene Nederlandse Spraakkunst* (1997). W. Haeseryn, K. Romijn, G. Geerts, J. de Rooij, M.C. Van den Toorn (red.). Groningen / Deurne: Martinus Nijhoff / Wolters Plantyn.
- Arnold, J.E., T. Wasow, A. Losongco & R. Ginstrom (2000). "Heaviness vs. newness: the effects of structural complexity and discourse status on constituent ordering." *Language* 76, 28-55.
- Baayen, R.H., R. Piepenbrock & H. Van Rijn (1993). *The CELEX lexical database*. Philadelphia: Linguistic Data Consortium.
- Barlow, M. & S. Kemmer (2000). *Usage-based models of language*. Stanford: CSLI.
- Beeken, J. & H. Smessaert (1996). "The syntax of Dutch connectives: some preliminary observations. *Leuvense bijdragen* 84/3, 375-396.

- Behaghel, O. (1930). *Die deutsche Sprache*. Leipzig: Freytag.
- Berg, B. Van den (1949). "De plaats van het hulpwerkwoord in de voltooide tijden in de Nederlandse bijzin". *Taal en tongval* 1, 155-165.
- Besten, H. Den & J.A. Edmonson (1983). "The verbal complex in Continental West Germanic". In: W. Abraham (red.), *On the formal syntax of the Westgermania*. *Linguistik Aktuell*. Amsterdamer Beiträge zur theoretischen und angewandten Linguistik, Band 3, 155-216.
- Besten, H. Den & H. Broekhuis (1989). "Woordvolgorde in de werkwoordelijke eindreeks". *Glott* 12/2, 79-137.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: CUP.
- Biber, D. (1995). *Dimensions of register variation*. Cambridge: CUP.
- Biber, D., S. Johansson, G. Leech & R. Quirk (1999). *Longman grammar of spoken and written English*. London: Longman.
- Bock, J.K. (1986). "Syntactic persistence in language production". *Cognitive psychology* 18, 355-387.
- Bock, K. & W. Levelt (1994). "Language production. Grammatical encoding". In: M.A. Gernsbacher (red.), *Handbook of psycholinguistics*. San Diego: Academic Press, 945-984.
- Bolinger, D. (1975). *Aspects of language*. New York: Harcourt Brace Jovanovich.
- Bolinger, D. (1977). *Meaning and form*. London: Longman.
- Booij, G. (1995). *The phonology of Dutch*. Oxford: Clarendon Press.
- Booij, G. & A. Van Santen (1998). *Morfologie: de woordstructuur van het Nederlands*. Amsterdam: Amsterdam University Press.
- Booij, G. (2002). "Constructional idioms, morphology, and the Dutch lexicon". *Journal of Germanic languages* 14/4, 301-329.
- Bouma, G. & G. Van Noord (1998). "Word order constraints on verb clusters in German and Dutch". In: E. Hinrichs, A. Kathol & T. Nakazawa (red.), *Complex predicates in Nonderivational Syntax*. San Diego: Academic Press, 43-72 (<http://odur.let.rug.nl/~vannoord/papers/complex/complex.html>).
- Brachin, P. & P. Vincent (1985). *The Dutch language. A survey*. Leiden: Brill.

- Braecke, C. (1986). “‘Zuidnederlandse’ volgorde in vier constructies: een zelfde analytische tendens?”. In: M. Devos & J. Taeldeman (red.), *Vruchten van z'n akker. Opstellen van (oud-) medewerkers en oud-studenten voor prof. V.F. Vanacker, hem aangeboden bij zijn afscheid van de Rijksuniversiteit Gent*. Gent: Seminarie voor Nederlandse Taalkunde en Vlaamse Dialectologie Rijksuniversiteit Gent, 33-45.
- Braecke, C. (1990). “Uit de tang' of ± prominent?” *Taal en Tongval* (themanummer 3: dialectsyntaxis), 125-134.
- Bree, C. Van (2000). “Taalbewustzijn, taalverandering en regionale taalvariatie. Een benadering”. *Taal en tongval* 52, 22-46.
- Briscoe, T. & J. Carroll (1997). “Automatic extraction of subcategorisation from corpora”. In: *Proceedings of the 5th ACL Conference on Applied Natural Language Processing*. Washington, 356-363.
- Bruch, R. (1973). *Précis populaire de grammaire luxembourgeoise*. Luxembourg: Goffin.
- Buchholz, S. (2002). *Memory-based Grammatical Relation Finding*. Proefschrift Universiteit Tilburg.
- Burridge, K. (1993). *Syntactic change in Germanic. Aspects of language change in Germanic with particular reference to Middle Dutch*. Amsterdam / Philadelphia: Benjamins.
- Bybee, J.L., P.J. Hopper (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.
- Canegem-Ardijns, I. Van (2004). *The use of prepositional objects of adjectives in brace- and extraposition constructions in Dutch*. Proefschrift K.U. Leuven.
- Chafe, W. (1988). “Punctuation and the prosody of written language”. *Written communication* 5/4, 396-426.
- Chomsky, N. (1972). *Language and the mind*. New York: Harcourt Brace Jovanovich.
- Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT.
- Claes, F. s.j. (1992). “Bijzinsvolgorde in oude Diesterse teksten”. *Taal en tongval* 44/1, 67-70.

- Clifton, C. Jr. & S.A. Duffy (2001). "Sentence and text comprehension: Roles of linguistic structure". *Annual Review of Psychology* 52, 167-196.
- Collier, R. & F.G. Droste (1982). *Fonetiek en fonologie*. Leuven / Amersfoort: Acco
- Coppen, P.-A. (1987). "Het Amazon-algoritme voor werkwoordelijke eindclusters". *Gramma* 11/3, 153-167.
- Coussé, E. (2003). "Volgordevariatie en herinterpretatie in de tweeledige werkwoordelijke eindgroep in de bijzin". *Taal en tongval* 55/2, 138-156.
- Cubber, W. De (1986). "'wat er den dag te voren is voorgevallen te Lokeren'. Bijdrage tot de studie van de dubbele werkwoordgroep". In: M. Devos & J. Taeldeman (red.), *Vruchten van z'n akker. Opstellen van (oud-)medewerkers en oud-studenten voor prof. V.F. Vanacker, hem aangeboden bij zijn afscheid van de Rijksuniversiteit Gent*. Gent: Seminarie voor Nederlandse Taalkunde en Vlaamse Dialectologie, Rijksuniversiteit Gent, 63-78.
- Dik, S.C. (1989). *The theory of functional grammar. 1: The structure of the clause*. Foris: Dordrecht.
- Dirven, R. (1974). *Kennismaking met de TGG. Transformationele en Generatieve Grammatica*. Leuven: Acco.
- Dittmer, A. (1982). "Verb (Zustandspartizip/Vorgangspartizip) oder Adjektiv?" *Kopenhager Beiträge zur Germanistischen Linguistik* 19, 48-84.
- Donselaar, W. Van (1995). *Effects of accentuation and given/new information on word processing*. PhD University of Utrecht.
- Dreumel, S. Van & Coppen, P.-A. (2003). "Surface analysis of the verbal cluster in Dutch". *Linguistics* 41/1, 51-81.
- Droste, F.G. & J.E. Joseph (1991). "Introduction: linguistic theory and grammatical description". In: F.G. Droste & J.E. Joseph (red.), *Linguistic theory and grammatical description*. Amsterdam / Philadelphia: Benjamins.
- Duden Grammatik der deutschen Gegenwartssprache* (1998). P. Eisenberg, H. Gelhaus, H. Henne, H. Sitta & H. Wellmann (red.). Mannheim / Leipzig / Wien / Zürich: Dudenverlag.
- Duinhoven, A.M. (1985). "De deelwoorden vroeger en nu". *Voortgang. Jaarboek voor de neerlandistiek* VI, 97-138.

- Duinhoven, A.M. (1987). "Hebben een koppelwerkwoord?". *Forum der letteren* 28/4 (themanummer *Algemene Nederlandse Spraakkunst*), 79-82.
- Duinhoven, A.M. (1997). *Middel nederlandse syntaxis synchroon en diachroon. De werkwoordgroep*. Groningen: Nijhoff.
- Duinhoven, A.M. (1998). "Concurrerende volgordepatronen in de werkwoordgroep". *Nederlandse taalkunde* 3/2, 96-119.
- Dunning, T. (1993). "Accurate methodes for the statistics of surprise and coincidence". *Computational linguistics* 19/1, 61-74.
- Ebert, R.P. (1978). *Historische Syntax des Deutschen*. Stuttgart : Metzler.
- Ebert, R.P. (1998). *Verbstellungswandel bei Jugendlichen, Frauen und Männern im 16. Jahrhundert*. Tübingen: Niemeyer.
- EGVD 1.4 – *Elektronische Grote Van Dale* (2004). D. Geeraerts (red.). Utrecht / Antwerpen: Van Dale Lexicografie.
- Everaert, M. (1993). "Vaste verbindingen (in woordenboeken)". *Spektator* 22/1, 3-27.
- Evers, A. (1975). *The transformational cycle in Dutch and German*. Dissertatie Universiteit Utrecht.
- Evers, A. (1992). "Begingroep-middengroep-eindgroep". In: H. Bennis & J.W. de Vries (red.), *De Binnenbouw van het Nederlands. Een bundel artikelen voor Piet Paardekooper*. Dordrecht: ICG, 55-72.
- Evers, A. (2003). "Verbal clusters and cluster creepers". In: P.A.M. Seuren & G. Kempen (red.), *Verb constructions in German and Dutch*. Amsterdam/Philadelphia: Benjamins, 43-89.
- Evert, S. & B. Krenn (2001). "Methods for the qualitative evaluation of lexical association measures". *Proceedings of the 39th annual meeting of the association for computational linguistics*, Toulouse, 188-195.
- Evert, S. & B. Krenn (2003). "Computational approaches to collocations". *European Summer School on Logic, Language, and Information (ESSLLI 2003)*, Vienna.
- Eynde, F. Van (2004). *Part of speech tagging en lemmatisering van het Corpus Gesproken Nederlands*. K.U.Leuven: Centrum voor Computerlinguïstiek.
- Fenk-Oczlon, G. (2001). "Familiarity, information flow, and linguistic form". In: J. Bybee & P. Hopper (red.), *Frequency and the emergence of linguistic structure*. Amsterdam / Philadelphia: Benjamins, 431-448.

- Friedl, J.E.F. (1998). *Mastering regular expressions*. Sebastopol: O'Reilly.
- Frings, T. & J. Vandenheuvel (1921). *Die südniederländischen Mundarten. Texte, Untersuchungen, Karten*. Marburg: Elwert'sche Verlagsbuchhandlung.
- Gebhardt, A. (1907). *Grammatik der Nürnburger Mundart*. Leipzig: Breitkopf & Härtel.
- Geeraerts, D., S. Grondelaers & P. Bakema (1994). *The structure of lexical variation. Meaning, naming, and context*. Berlin: Mouton de Gruyter.
- Geeraerts, D., S. Grondelaers & D. Speelman (1999). *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen*. Amsterdam: KNAW.
- Geeraerts, D. (2003). "Decontextualizing and recontextualizing tendencies in 20th-century linguistics and literary theory". In: E. Mengel, H.-J. Schmid & M. Steppat (red.), *Anglistentag 2002 Bayreuth*. Trier: Wissenschaftlicher Verlag, 369-379.
- Geeraerts, D. (2004). "Corpuslinguïstiek in context". Doctoraatsseminarie voor het departement Linguïstiek K.U.Leuven.
- Gerritsen, M. (1978). "De opkomst van SOV-patronen in het Nederlands in verband met woordvolgordeveranderingen in de Germaanse talen". In: J.G. Kooij (red.), *Aspekten van woordvolgorde in het Nederlands*. Leiden: Publikaties van de vakgroep Nederlandse taal- en letterkunde, 3-39.
- Gerritsen, M. (1991). *Atlas van de Nederlandse dialectsyntaxis (AND)*. Publikaties van het P.J. Meertensinstituut voor dialectologie, volkskunde en naamkunde van de Koninklijke Nederlandse Akademie van Wetenschappen 17. Amsterdam: KNAW.
- Gerritsen, M. (1992). "Internal and external factors in the stabilization of verb-last order in Dutch infinitive clauses". In: M. Gerritsen & D. Stein (red.), *Internal and external factors in syntactic change*. Berlin / New York: De Gruyter, 355-394.
- Gibson, E. (1998). "Linguistic complexity: locality of syntactic dependencies". *Cognition* 68, 1-76.

- Givón, T. (1988). "The Pragmatics of Word-Order: Predictability, Importance and Attention." In: M. Hammond, E.A. Moravcsik & J.R. Wirth (red.), *Studies in syntactic typology*. Amsterdam: Benjamins, 243-284.
- Greenberg, J.H. (1973). *Some universals of grammar with particular reference to the order of meaningful elements*. In: J.H. Greenberg (red.), *Universals of language*. Cambridge & London: MIT, 73-113.
- Gries, S.Th. & A. Stefanowitsch (2004). "Extending collocation analysis. A corpus-based perspective on 'alternations'". *International journal of corpus linguistics* 9/1, 97-129.
- Grishmann, R., C. Macleod & A. Meyers (1994). "Complex Syntax: Building a computational lexicon." *COLING 94* [manuscript].
- Grondelaers, S. (2000). *De distributie van niet-anaforsch er buiten de eerste zinsplaats. Sociolexicologische, functionele en psycholinguïstische aspecten van er's status als presentatief signaal*. Proefschrift K.U. Leuven.
- Grondelaers, S., K. Deygers, H. Van Aken, V. Van Den Heede & D. Speelman (2000). "Het CONDIV-corpus geschreven Nederlands". *Nederlandse taalkunde* 5/4, 356-363.
- Grondelaers, S., D. Speelman & A. Carbonez (2001). "Regionale variatie in de postverbale distributie van presentatief er". *Neerlandistiek.nl* 01.04. <<http://www.neerlandistiek.nl/01.04/>>
- Haan, G.J. De, G.A.T. Koefoed & A.L. des Tombe (1975). *Basiskursus algemene taalwetenschap*. Assen / Amsterdam: Van Gorcum.
- Haas, W. De & M. Trommelen (1993). *Morfologisch handboek van het Nederlands: een overzicht van de woordvorming*. 's Gravenhage: SDU.
- Haegeman, L. & H. Van Riemsdijk (1986). "Verb projection raising, scope and the typology of rules affecting verbs". *Linguistic inquiry* 17, 417-466.
- Haeseryn, W. (1990). *Syntactische normen in het Nederlands. Een empirisch onderzoek naar volgordevariatie in de werkwoordelijke eindgroep*. Proefschrift K.U. Nijmegen.
- Haeseryn, W. (1996). "Grammaticale verschillen tussen het Nederlands in België en het Nederlands in Nederland: een poging tot inventarisatie". In: R. Van Hout & J.

- Kruijssen (red.), *Taalvariatie. Toonzettingen en modulaties op een thema*. Dordrecht: Foris, 109-126.
- Haeseryn, W. (1999). "Normatieve studies". In: W. Smedts & P.C. Paardekooper (red.), *De Nederlandse taalkunde in kaart*. Leuven: Acco, 237-247.
- Harrell, F.E. jr. (2001). *Regression modelling strategies. With applications to linear models, logistic regression, and survival analysis*. New York: Springer.
- Hart, J. 't, R. Collier & A. Cohen (1990). *A perceptual study of intonation. An experimental-phonetic approach to speech melody*. Cambridge e.a.: Cambridge University Press.
- Hartsuiker, R.J., H.H.J. Kolk & P. Huiskamp (1999). "Priming word order in sentence production". *The quarterly journal of experimental psychology* 52A/1, 129-147.
- Hartsuiker, R.J. & C. Westenberg (2000). "Word order priming in written and spoken sentence production". *Cognition* 75, B27-B39.
- Hawkins, J.A. (1994). *A performance theory of order and constituency*. Cambridge: CUP.
- Hawkins, J.A. (2000). "The relative order of prepositional phrases in English: Going beyond Manner-Place-Time". *Language Variation and Change* 11, 231-266.
- Heidbuchel, H. (1967). *Hoe zeg en schrijf ik het?* Hasselt: HeideLand.
- Helbig, G. & W. Schenkel (1991). *Wörterbuch zur Valenz und Distribution deutscher Verben*. Tübingen: Niemeyer.
- Hertog, C.H. den (1972-1973). *Nederlandse spraakkunst*. Amsterdam: Versluys (bewerkt door H. Hulshof).
- Hodler, W. (1969). *Berndeutsche Syntax*. Berlin: Francke Verlag.
- Hoop, H. De & P. Smabers (1987). "Rood of groen? Een normenonderzoek naar enkele stijlverschillen tussen gesproken en geschreven taal". *De nieuwe taalgids* 80/4, 287-302.
- Hopper, P.J. & E.C. Traugott (2003). *Grammaticalization*. Cambridge: Cambridge University Press.
- Horst, J.M. Van der (1981). *Kleine Middelnederlandse syntaxis*. Amsterdam: Huis aan de drie grachten

- Horst, J.M. Van der (1993/1994). “Rode en groene volgorde en analytische taalkunde”. *Voortgang. Jaarboek voor de neerlandistiek* 14, 187-197.
- Horst, J.M. Van der (2000). “Rood licht voor groen”. *Nederlands van nu* 48/5, 92-94.
- Horst, J.M. Van der (2002). “Geschiedenis van een taalnorm: de doorbroken werkwoordelijke eindgroep (deel I)”. In: R. Willemys (red.), *De taal in Vlaanderen in de 19^{de} eeuw. Historisch-sociolinguïstische onderzoeken*. Verslagen en mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde. Gent: KANTL, 529-556.
- Horst, J. Van der & K. Van der Horst (1999). *Geschiedenis van het Nederlands in de twintigste eeuw*. Den Haag / Antwerpen: Sdu / Standaard.
- Hotzenköcherle, R. (red.) (1975). *Sprachatlas der deutschen Schweiz*. Bern: Francke.
- Jackendoff, R. (1977). *X-bar syntax: a study of phrase structure*. Cambridge: MIT Press.
- Jacobs, J. (1994). *Kontra Valenz*. Trier: Wissenschaftlicher Verlag.
- Jansen, F. (1979). “On tracing conditioning factors of movement rules: extraposition of PP in spoken Dutch”. In: M. Van de Velde & W. Vandeweghe (red.), *Sprachstruktur, Individuum und Gesellschaft*. Tübingen: Niemeyer, 83-92.
- Jansen, F. (1990). “Tangconstructies en geen reacties daarop”. *Taal en tongval* (themanummer 3: dialectsyntaxis), 135-151.
- Jansen, F. & R. Wijnands (2004). “Doorkruisingen van het Links-RechtsPrincipe”. *Neerlandistiek.nl* 04.01. <<http://www.neerlandistiek.nl/04.01/>>
- Janssen, T.A.J.M. (1985). “De constructie hebben/zijn + (‘)voltooid deelwoord(‘)”. *Voortgang. Jaarboek voor de neerlandistiek* VI, 49-84.
- Janssen, T.A.J.M. (1986). “Het voltooid deelwoord”. *Glottol* 9/1-2, 57-78.
- Kaplan, R.M. & A. Zaenen (2003). “West-Germanic Verb Clusters in LFG”. In: P.A.M. Seuren & G. Kempen (red.), *Verb constructions in German and Dutch*. Amsterdam/Philadelphia: Benjamins, 127-150.
- Katwijk, A. Van (1974). *Accentuation in Dutch. An experimental linguistic study*. Proefschrift Universiteit Utrecht.
- Kempen, G. & K. Harbusch (2003). “Dutch and German verb constructions in Performance Grammar”. In: P.A.M. Seuren & G. Kempen (red.), *Verb*

- constructions in German and Dutch*. Amsterdam/Philadelphia: Benjamins , 185-221.
- Kerstens, J.G. & A.N. Sturm (1979). "Over problemen met intuities". *De nieuwe taalgids* 72/5, 449-465.
- Kerstens, J., E. Ruys, M. Trommelen & F. Weerman (1997). *Plato's probleem. Een inleiding in de generatieve taalkunde*. Bussum: Coutinho.
- Kilgarriff, A. (2001). "Comparing corpora". *International journal of corpus linguistics* 6/1, 97-133.
- Klein, H. (1998). *Adverbs of degree in Dutch and related languages*. Amsterdam: Benjamins.
- Klein, M. (1999). Syntaxis. In: W. Smedts & P.C. Paardekooper (red.), *De Nederlandse taalkunde in kaart*. Leuven: Acco.
- Klooster, W. (2001). *Grammatica van het hedendaags Nederlands. Een volledig overzicht*. Den Haag: Sdu.
- Knops, U. & R. Van Hout (1988). "Language attitudes in the Dutch language area: an introduction. In: U. Knops & R. van Hout (red.), *Language attitudes in the Dutch language area*. Dordrecht / Providence: Foris.
- Ko, K. (1996). "Structural characteristics of computer-mediated language: a comparative analysis of Interchange discourse". *Electronic journal of communication* 6.
<<http://ublib.buffalo.edu/libraries/eresources/ejournals/records/ejc.html>>
- Koelmans, L. (1965). "Iets over de woordorde bij samengestelde predikaten in het Nederlands". *De nieuwe taalgids* 58, 156-165.
- Koelmans, L. (2001) *Het Nederlands van Michiel de Ruyter: morfologie, woordvorming, syntaxis*. Assen: Van Gorcum.
- Labov, W. (1972). "Some principles of linguistic methodology". *Language in society* 1, 97-120.
- Ladd, D.R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Lamers, H.A.J.M. (1997). Handleiding voor beleidsteksten: een handleiding om beleidsnota's, beleidsbrieven, jaarverslagen en notulen op te stellen. Bussum: Coutinho.

- Langacker, R. (1987). *Foundations of cognitive grammar. Theoretical prerequisites*. Stanford: Stanford University Press.
- Langendonck, W. Van (2003). "Dependenz. Grundlagen und Grundfrage. Dependency. Basic principles and basic issues". In: V. Agel, L.M. Eichinger, H.-W. Eroms, P. Hellwig, H.J. Heringer & H. Lobin, *Dependenz und Valenz. Dependency and valency. Ein internationales Handbuch der zeitgenössischen Forschung. An international handbook of contemporary research*. Berlin / New York: De Gruyter.
- Lenz, B. (1993). "Probleme der Kategorisierung deutscher Partizipien". *Zeitschrift für Sprachwissenschaft* 12/1, 39-76.
- Leuvensteijn, J.A. Van (1987). "De werkwoordelijke eindgroep in Gouds taalgebruik uit de 14^{de} en de 16^{de} eeuw". *Taal en tongval* 39, 1-20.
- Levin, B. & M. Rappaport (1986). "The formation of adjectival passives". *Linguistic inquiry* 17/4, 623-661.
- Lubbe, H.F.A. Van der (1965). *Woordvolgorde in het Nederlands. Een synchrone structurele beschouwing*. Assen: Van Gorcum.
- Lubbe, H.F.A. Van der (1980). "De waarden van de vrijwillige armoede". *De nieuwe taalgids* 73/3, 234-238.
- McDonald, J., K. Bock, & M.H. Kelly (1993). "Word and world order: semantic, phonological, and metrical determinants of serial position". *Cognitive Psychology* 25/2, 188-230.
- Marynissen, A. (1999). "...allen dengen die dese letteren sien selen / selen sien ende horen lesen ...Over volgordevariatie in de werkwoordelijke eindgroep in de Middelnederlandse bijzin". *Taal en tongval*, themanummer 12, 136-158.
- Meersman, A. De (1990). "Bijzinsvolgorde in tweeledige ww-groepen. Een verkennende historische schets". *Taal en tongval*, themanummer 3, 152-185.
- Meeussen, A.E. & V.F. Vanacker (1951). "De dubbele werkwoordgroep". *De nieuwe taalgids* 44, 36-50.
- Mendoza-Denton, N., J. Hay & S. Jannedy (2003). "Probabilistic sociolinguistics: beyond variable rules". In: R. Bod, J. Hay & S. Jannedy (red.), *Probabilistic linguistics*. Cambridge: MIT.

- Meyers, A., C. Macleod & R. Grishman (1996). "Standardization of the Complement/Adjunct distinction". *Proteus Project Memorandum 64*, New York: New York University [manuscript].
- Michels, L.C. (1957). "Op de grens van copula en hulpwerkwoord". *Taal en tongval* XI, 106-112.
- Milroy, L. & M. Gordon (2003). *Sociolinguistics. Method and interpretation*. Oxford: Blackwell.
- Mitchell, D.C., F. Cuetos, M.M.B. Corley & M. Brysbaert (1995). "Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records". *Journal of Psycholinguistic Research* 24/6, 469-488.
- Moon, R. (1998). *Fixed expressions and idioms in English. A corpus-based approach*. Oxford: Clarendon Press.
- Nooteboom, S. (1997). "The prosody of speech: melody and rhythm". In: W.J. Hardcastle & J. Laver (red.), *The handbook of phonetic sciences*. Oxford / Cambridge: Blackwell, 640-673.
- Oostdijk, N. (2000). "Het Corpus Gesproken Nederlands". *Nederlandse taalkunde* 5, 280-284.
- Overdiep, G.S. (1937). *Stilistische grammatica van het moderne Nederlandsch*. Zwolle: Tjeenk Willink.
- Overdiep, G.S. (1949). *Stilistische grammatica van het moderne Nederlandsch*. Zwolle: Tjeenk Willink.
- Overduin, B. (1986). *Rapporteren. Het schrijven van rapporten, nota's, scripties en artikelen*. Utrecht: Aula.
- Paolillo, J.C. (2002). *Analyzing linguistic variation: statistical models and methods*. Stanford: CSLI.
- Pardoen, J.A. (1991). "De interpretatie van zinnen met de rode of de groene volgorde". *Forum der letteren* 32/1, 1-22.
- Pardoen, J.A. (1998). *Interpretatiestructuur. Een onderzoek naar de relatie tussen woordvolgorde en zinsbetekenis in het Nederlands*. Amsterdam / Münster: Stichting Neerlandistiek VU / Nodus.

- Pauwels, A. (1953). *De plaats van hulpwerkwoord verleden deelwoord en infinitief in de Nederlandse bijzin*. Werken uitgegeven door de Koninklijke Commissie voor Toponymie en Dialectologie 7. Leuven: Symons.
- Pickering, M.J., H.P. Branigan, A.A. Cleland & A.J. Stewart (2000). "Activation of syntactic information during language production". *Journal of psycholinguistic research* 29/2, 205-216.
- Pintzuk, S. & E. Haeberli (2004). "Reconsidering the status of verb (projection) raising in Old English". Lezing gehouden op NWAV 33, University of Michigan, Ann Arbor.
- Pollard, C.J. & I.A. Sag (1987). *Information-based syntax and semantics. Volume I: fundamentals*. Chicago: Chicago University Press.
- Quak, A. & J.M. Van der Horst (2002). *Inleiding Oudnederlands*. Leuven: UP.
- R Development Core Team (2003). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <<http://www.R-project.org>>
- Renkema, J. (1995). *Schrijfwijzer*. 's Gravenhage: Sdu.
- Rietveld, T. & R. van Hout (1993). *Statistical techniques for the study of language and language behaviour*. Berlin / New York: Mouton.
- Rooij, J. De (1986). "Waarom het naamwoordelijk gezegde groen blijft". *Onze taal* 55/11, 142-144.
- Sassen, A. (1963). "Endogeen en exogeen taalgebruik". *De nieuwe taalgids* 56/1, 10-21.
- Scholten, A., A. Evers & M. Klein (1981). *Inleiding in de transformationeel-generatieve taaltheorie*. Groningen: Wolters-Noordhoff.S
- Schutter, G. De (1964). "De dubbele werkwoordgroep in het Nederlands". *Studia germanica gandensia* VI, 45-84.
- Schutter, G. De (1967). "Principes van de woordvolgorde in de Nederlandse zin". *Handelingen van het XXVIe Vlaams filologencongres*, 47-55.
- Schutter, G. De (1976). "De bouw van de Nederlandse zin. Beschrijving en voorstel tot beregeling". *Verslagen en mededelingen van de Koninklijke Academie voor Taal en Letterkunde*. Gent: KANTL.

- Schutter, G. De (1991). "Syntactische verklaringen voor werkwoordvolgordes". *De nieuwe taalgids* 84/4, 312-319.
- Schutter, G. De (1996). "De volgorde in tweeledige werkwoordelijke eindgroepen met voltooid deelwoord in spreek- en schrijftaal". *Nederlandse taalkunde* 1, 207-220.
- Schutter, G. De (2003). "De zinsbouw in het vroege Middelnederlands. Een case study". *Taal en tongval* 55/1, 41-67.
- Schütze, C.T. (1996). *The empirical base of linguistics. Grammaticality judgments and linguistic methodology*. Chicago / London: Chicago University Press.
- Selkirk, E. (1995). "Sentence prosody: intonation, stress, and phrasing". In: J.A. Goldsmith (red.), *The handbook of phonological theory*. Cambridge / Oxford: Blackwell, 551-569.
- Seuren, P.A.M. (2003). "Verb clusters and branching directionality in German and Dutch". In: P.A.M. Seuren & G. Kempen (red.), *Verb constructions in German and Dutch*. Amsterdam/Philadelphia: Benjamins, 247-296.
- Shannon, T.F. (1995). "Extrapolation of NP complements in Dutch and German: an empirical comparison". In: T.F. Shannon & J.P. Snapper (red.), *The Berkeley conference on Dutch linguistics 1993. Dutch linguistics in a changing Europe*. Lanham/New York/London: University Press of America, 87-116.
- Siewierska, A. (1993). "Syntactic weight vs. information structure and word order variation in Polish". *Journal of Linguistics* 29, 233-265.
- Simpson, G.B. (1994). "Context and the processing of ambiguous words". In: M.A. Gernsbacher (red.), *Handbook of psycholinguistics*. San Diego: Academic Press, 359-374.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Smedts, W. & W. Van Belle (1994). *Taalboek Nederlands*. Kapellen: DNB Pelckmans.
- Somers, H. (1984). "On the validity of the complement-adjunct distinction in valency grammar". *Linguistics* 22, 507-530.

- Sommerfeldt, K.-E. (1988). "Zur Wortartklassifizierung des Deutschen— Untersuchungen zur Wortartzugehörigkeit partizipialer Bildungen". *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 41/2, 221-229.
- Speelman, D. (1997). *Abundantia Verborum. A computer tool for carrying out corpus-based linguistic case studies*. Proefschrift K.U. Leuven. <http://wwwling.arts.kuleuven.ac.be/genling/abundant/>
- Sperschneider, H. (1959). *Studien zur Syntax der Mundarten im östlichen Thüringer Wald*. Deutsche Dialektgeographie 54. Marburg: Elwert.
- Stoett, F.A. (1909). *Middelnederlandsche spraakkunst. Syntaxis*. 's Gravenhage: Martinus Nijhoff.
- Storrer, A. (2003). "Ergänzungen und Angaben". In: V. Agel, L.M. Eichinger, H.-W. Eroms, P. Hellwig, H.J. Heringer & H. Lobin (red.), *Dependenz und Valenz. Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin: De Gruyter.
- Stroobants, K. (1997). "Noord/Zuid-verschillen in de taal van de reclame". *Taal en tongval* 49, 140-151.
- Stroop, J. (1970). "Systeem in gesproken werkwoordsgroepen". *Taal en tongval* 22, 128-146.
- Sturm, A. (1990). "Een syntactische verklaring voor de rode en groene volgorde". *De nieuwe taalgids* 83/5, 385-404.
- Sturm, A. & J. Kerstens (1981). "Over de feiten van de generatieve taalkunde en de waarneming ervan". *De nieuwe taalgids* 74/2, 149-175.
- Sutter, G. De, D. Speelman, D. Geeraerts & S. Grondelaers (2003). "Woordvolgordevariatie in werkwoordelijke eindgroepen. Een statistische evaluatie van zes factoren". In: T. Koole, J. Nortier & B. Tahitu (red.), *Artikelen van de vierde sociolinguïstische conferentie*. Delft: Eburon, 111-121.
- Sutter, G. De, D. Speelman & D. Geeraerts (2005). "Regionale en stilistische effecten op de woordvolgorde in werkwoordelijke eindgroepen". *Nederlandse taalkunde* 10/2.
- Swerts, M. (1998). "Ritme als verklarende factor voor de keuze tussen groene en rode werkwoordvolgorde in het Nederlands". *Nederlandse taalkunde* 3/4, 299-308.
- Swerts, M. & R. Collier (2000). "Preferentieoordelen van luisteraars voor uitingen met groene of rode werkwoordvolgorde". In: S. Gillis, J. Nuyts & J. Taeldeman

- (red.), *Met taal om de tuin geleid. Een bundel opstellen voor Georges De Schutter ter gelegenheid van zijn pre-emeritaat*. Antwerpen: UIA, 311-316.
- Taldeman, J. (1992). "Welk Nederlands voor de Vlamingen?" *Nederlands van nu* 40/2, 33-52.
- Takada, H. (1998). *Grammatik und Sprachwirklichkeit von 1640-1700: zur Rolle deutscher Grammatiker im schriftsprachlichen Ausgleichsprozess*. Tübingen: Niemeyer.
- Tomasello, M. (1998). *The new psychology of language. Cognitive and functional approaches to language structure*. New Jersey: Erlbaum.
- Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge & London: Harvard University Press.
- Toorn, M.C. Van den, J.A. Van Pijnenburg, W.J.J. Leuvensteijn & J.M. Van der Horst (1997). *Geschiedenis van de Nederlandse taal*. Amsterdam: AUP.
- Vanacker, V.F. (1963). *Syntaxis van gesproken taal te Aalst en in het Land van Aalst in de XVde, de XVIde en de XVIIde eeuw*. Brussel: Belgisch inter-universitair centrum voor neerlandistiek.
- Van Dale groot woordenboek van hedendaags Nederlands* (1991). P.G.J. van Sterkenburg & G.E. Booij (red.). Utrecht: Van Dale Lexicografie.
- Verhagen, A. (1986). *Linguistic theory and the function of word order in Dutch. A study on interpretive aspects of the order of adverbials and noun phrases*. Dordrecht: ICG.
- Verhagen, A. (1992). "Praxis of linguistics: passives in Dutch". *Cognitive linguistics* 3/3, 301-342.
- Verhasselt, J. (1961). "Verschillen tussen Noord en Zuid inzake de volgorde hulpwerkwoord – hoofdwerkwoord". *Taal en tongval* 13/2-3, 153-157.
- Verhoeven, G. (1981a). "Toch problemen met intuïties". *De nieuwe taalgids* 74/1, 23-28.
- Verhoeven, G. (1981b). "Het waarnemen van de taalkunde". *De nieuwe taalgids* 74/4, 354-361.
- Versteeg, T.J.M. (1969). "'Zijn' + voltooid deelwoord". *De nieuwe taalgids* 62, 422-424.

- Verstraten, L.P. (1992). *Vaste verbindingen: een lexicologische studie vanuit cognitief-semantic perspectief naar fraseologismen in het Nederlands*. Proefschrift Universiteit Leiden.
- Vooyo, C.G.N. De & M. Schönfeld (1957). *Nederlandse spraakkunst*. Groningen: Wolters.
- Vos, R. (1968). "Over het werkwoordelijk en naamwoordelijk gezegde". *De nieuwe taalgids* 61, 217-231.
- Vries, J. De, F. de Tollenaere & M. Hogenhout-Mulder (1971). *Nederlands etymologisch woordenboek*. Leiden: Brill.
- Vries, M. De & L.A. Te Winkel (1882-1998). *Woordenboek der Nederlandsche taal*. 's Gravenhage: Nijhoff.
- Wal, M. Van der & C. Van Bree (1994). *Geschiedenis van het Nederlands*. Utrecht: Spectrum.
- Walter, M. & K.P. Gietzen (red.) (2003). *Mönchengladbacher Mundartwörterbuch von Johannes Noever*. Unsere Heimat in Wort und Bild 13. Mönchengladbach: Heimat- und Geschichtsverein Mönchengladbach.
- Wasow, T. (1977). "Transformations and the lexicon". In: P.W. Culicover, T. Wasow & A. Akmajian, *Formal syntax*. New York / San Francisco / London: Academic Press.
- Wasow, T. (2002). *Postverbal behavior*. Stanford: CSLI.
- Wild, K. (1997). Zur Verbstellung in den schwäbischen Mundarten Südungarns. In: A. Ruoff & P. Löffelad (red.), *Syntax und Stilistik der Alltagssprache*. Tübingen: Niemeyer, 145-154.
- Zuckerman, S. (2001). *The acquisition of 'optional' movement*. Groningen: Groningen Dissertations in Linguistics 34.
- Zwart, J.-W. (1996). "Verb clusters in Continental West Germanic dialects". In: J.R. Black & V. Motapanyane (red.), *Microparametric syntax and dialect variation*. Amsterdam / Philadelphia: Benjamins, 229-258.

